DOCUMENT RESUME

ED 181 421 CS 005 198

AUTHOR	Thorndyke, Perry W.; Hayes-Roth, Barbara
TITLE	Human Processing of Knowledge from Texts: Acquisition, Integration, and Reasoning.
INSTITUTION	Rand Corp., Santa Monica, Calif.
SPONS AGENCY	Advanced Research Projects Agency (DOD), Washington, D.C.
REPORT NO	ISBN-08330-0149-3; R-2256-ARPA
PUB DATE	Jun 79
CONTRACT	DAHC15-73-C-0181
NOTE	166p.
AVAILABLE FROM	The Rand Corporation, Santa Monica, CA 90406 ($7.00)
EDRS PRICE	MF01/PC07 Plus Postage.
DESCRIPTORS	Advance Organizers; *Cognitive Processes; *Directed Reading Activity; Higher Education; *Learning Processes; *Reading Comprehension; Reading Materials; *Reading Research; Research Projects; *Retention; Retention Studies

ABSTRACT

This report documents a series of studies on how undergraduate students learn from and reason with textual information. The studies described were undertaken to produce models that could serve as the basis for designing computer systems capable of structuring and presenting text material in optimal formats. Divided into sections, the report presents five sets of studies evaluating particular processing models and optimization techniques to improve reading comprehension and retention, including the use of schemata in the acquisition and transfer of knowledge, knowledge acquisition from newspaper stories, the integration of knowledge from text, configural effects in memory, and text annotation as a technique for facilitating knowledge integration. The last section of the report discusses ten conclusions, among which are that presenting new information in well-learned structural organizations can facilitate learning that information, that eliminating redundancy and irrelevant commentary from newspaper stories facilitates assimilation and retention of important facts, and that annotating texts with references to related facts that have occurred in previous texts facilitates general inferential reasoning from the texts. (AEA)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED181421

ARPA Order No.: 189-1
9D20 Cybernetics Technology

U S DEPARTMENT OF HEALTH.
EDUCATION & WELFARE
NATIONAL INSTITUTE OF
EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

R-2256-ARPA
June 1979

Human Processing of Knowledge from Texts: Acquisition, Integration, and Reasoning

Perry W. Thorndyke, Barbara Hayes-Roth

with the assistance of Susan Knobel, James Miller, and Carol Walker

A Report prepared for

DEFENSE ADVANCED RESEARCH PROJECTS AGENCY

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
Rand Corp.

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

Rand
SANTA MONICA, CA. 90406

CS005198

The research described in this report was sponsored by the Defense Advanced Research Projects Agency under Contract No. DAHC15-73-C-0181.

Library of Congress Cataloging in Publication Data

Thorndyke, Perry W
Human processing of knowledge from texts.

([Report] - Rand Corporation ; R-2256-ARPA)
Bibliography: p.
1. Human information processing. 2. Learning, Psychology of. 3. Reasoning (Psychology) I. Hayes-Roth, Barbara, joint author. II. United States. Defense Advanced Research Projects Agency. III. Title. IV. Series: Rand Corporation. Rand report ; R-2256-ARPA.
AS36.R3 R-2256 [BF455] 081s [153.4] 79-17764
ISBN 0-8330-0149-3

The Rand Publications Series: The Report is the principal publication documenting and transmitting Rand's major research findings and final research results. The Rand Note reports other outputs of sponsored research for general distribution. Publications of The Rand Corporation do not necessarily reflect the opinions or policies of the sponsors of Rand research.

Published by The Rand Corporation

PREFACE

Two fundamental components of the decisionmaking process are the acquisition of new knowledge and the retrieval of knowledge from memory. The knowledge available to a decisionmaker is frequently expressed in textual documents that are voluminous and poorly organized. Decisionmakers in command and control situations, both currently and in the future, will become increasingly dependent on textual knowledge bases that are resident in computers. The quality of decisions will undoubtedly depend on the quality of the information available to the decisionmaker. Thus it is important to optimize both the amount of information he can assimilate from a text and his ability to retrieve that information when needed. To improve his ability to retain and use acquired knowledge, it is necessary to understand the relationship between the information presented in texts and the representation of knowledge in human memory. It should be possible to adapt the format of the presented text to facilitate the reader's assimilation and utilization of the information. This might be accomplished by filtering and reorganizing the to-be-learned material. Developing technologies in electronic data bases and computer-controlled communications make this filtering process a real possibility for decisionmakers who routinely receive information on-line.

This report details the results of a one-year study, completed in November 1977, supported by the Cybernetics Technology Office (CTO) of the Defense Advanced Research Projects Agency (ARPA). The findings, which were provided to ARPA at that time in informal documentation, have been prepared here in report form, using Rand Corporation funds, for distribution to a wider audience.

The objective of the research was to develop models of how humans store, organize, and retrieve in memory information obtained from reading texts. These models, derived empirically from a series of psychological experiments, might serve as the basis for the design of computer systems capable of structuring and presenting texts in optimal formats. A system incorporating such principles of human learning into its text-handling facilities would prove to be a useful memory aid for military commanders, intelligence analysts, or other high-level decisionmakers who depend on large computer data bases of knowledge. The research reported here has been directed toward the development of human information processing models rather than a demonstrable computer system. However, the results of this study may have direct applicability to the construction of a system incorporating techniques for formatting texts into optimal organizations.

SUMMARY

This report documents a series of studies of how people learn from and reason with information contained in texts. The experiments reported here are based on the premise that readers typically derive their knowledge from several source texts. The reader's problem, then, is not only to acquire individual facts but to organize related facts obtained from diverse sources.

The research has been motivated by four goals: (1) to elucidate the process by which knowledge is acquired from a textual document, (2) to specify how the acquired knowledge is represented in memory, (3) to identify how and when related information is integrated in memory, and (4) to discover techniques for the facilitation of learning and reasoning with textual information. Two task domains were chosen in which to study the structures and processes underlying human performance with texts. The first task domain was simple ACQUISITION, in which the reader attempted to learn and retain as much knowledge from the text as possible. The second task domain was INFERENTIAL REASONING, in which the reader had to organize a set of facts from the text in order to generate or verify a conclusion following from the facts. For both domains, the approach taken was to model the representation of knowledge in memory and the processes required to perform the task, and then to design text presentation formats that facilitated both the transfer of knowledge from texts into these memory structures and the performance of inferential reasoning tasks.

Five sets of studies evaluated particular processing models and optimization techniques. In Study 1, the acquisition of new knowledge that conformed to a previously learned structure, or schema, was investigated. Memory for text information was generally facilitated by prior training with the structure in which the text was embedded, particularly when there was no confusion among facts from different texts sharing a common organization.

In Study 2, information from newspaper stories was restructured into different formats in an attempt to improve memory for the stories. These formats included a condensed version of the news story (with all redundant and superfluous information deleted), a narrative structure based on the temporal order of events, a topical structure organized according to the primary concepts of the passage, and an outline structure that reduced the text to key phrases spatially arranged on the page in outline form. All forms of restructuring produced significant improvements in recall of the information contained in the passage.

Study 3 investigated the integration of related but separately acquired facts in memory either for apprehension of a complex idea or for performing inferential reasoning. Integration of separate facts into a unified memory structure was more likely to occur if related facts occurred in close temporal proximity and if the facts cued each other with similar wordings, rather than with paraphrased wordings. The integration of facts significantly improved the ability to recall related information and perform syllogistic reasoning.

Techniques for improving a learner's ability to organize diverse information for inferencing were investigated in Studies 4 and 5. In Study 4, subjects verified inferences based on information from two distinct texts. Performance was better when subjects had attempted to commit the texts to memory than when they were allowed to inspect the texts freely during the reasoning test. In Study 5, methods of annotating texts to facilitate integration and inferencing were investigated.

The results from these studies are presented in the context of a set of models for knowledge representation and processing. Based on these models and the obtained experimental results, a set of principles for improving human learning from texts emerged. To the extent that our subject and materials samples were representative of the population of readers and potential texts, these principles are descriptive of human text-processing characteristics. They may be summarized as follows:

1. Presentation of new information in well-learned structural organizations can facilitate learning of the new information.
2. Presentation in close temporal proximity of large numbers of facts (i.e., five or more) sharing a common structural framework interferes with learning.
3. Temporal separation in presentation of interfering facts can limit interfering effects.
4. Elimination of redundancy and irrelevant commentary from newspaper stories facilitates assimilation and retention of important facts.
5. Text organizations that place complementary facts in close proximity improve integration of those facts.
6. Wording complementary texts as similarly as possible improves integration of complementary facts that occur in separate texts.
7. Wording related texts as similarly as possible improves inferential reasoning based on facts within the texts.
8. Reasoning from memory about carefully studied texts is more accurate than reasoning based on inspection of less familiar texts.
9. Knowledge of the information contained in texts is improved by studying to learn the texts rather than using the texts to perform inferencing.
10. Annotating texts with references to related facts that have occurred in previous texts facilitates general inferential reasoning from the texts.

ACKNOWLEDGMENTS

We wish to express our gratitude to the many individuals who contributed to this study. We are indebted to Cathleen Stasz, Elizabeth Weinberger, and Rick Yekovich for their assistance in testing subjects, scoring protocols, and analyzing data. We are also indebted to Frederick Hayes-Roth for helpful comments throughout the course of the research.

CONTENTS

8

PREFACE iii

SUMMARY v

ACKNOWLEDGMENTS ix

Section

I. INTRODUCTION 1

II. THE USE OF SCHEMATA IN THE ACQUISITION AND TRANSFER OF KNOWLEDGE 6
  Experiment 1 13
  Experiment 2 30
  General Discussion 37

III. KNOWLEDGE ACQUISITION FROM NEWSPAPER STORIES 42
  Experiment 3 45
  Experiment 4 54
  General Discussion 59

IV. INTEGRATION OF KNOWLEDGE FROM TEXT 63
  Experiment 5 70
  Experiment 6 78
  Experiment 7 84
  General Discussion 92

V. CONFIGURAL EFFECTS IN MEMORY 95
  Modeling Integrative Inferential Reasoning 97
  Experiment 8 101
  Practical Implications 128

VI. TEXT ANNOTATION: A TECHNIQUE FOR FACILITATING KNOWLEDGE INTEGRATION 132
  Experiment 9 135
  General Discussion 142

VII. CONCLUSIONS 144

REFERENCES 151

I. INTRODUCTION

An information glut confronts most decisionmakers. The information they must assimilate from textual documents is typically voluminous, poorly organized and presented, and informationally sparse. It is often necessary to make rapid decisions based on an overload of available information, and the decisionmaker can rarely invest the time necessary to rehearse the new material, integrate the new facts with existing knowledge, or otherwise improve his memory for the information. Presumably, the quality of decisions depends on the quality of the information that is utilized in arriving at those decisions. Therefore, optimizing a decisionmaker's ability to acquire and use information would have beneficial consequences for the decisionmaking process.

The objective of the research reported here was to develop models of how people store, organize, and retrieve in memory information obtained from reading texts. Throughout the course of the research, it was assumed that to improve the decisionmaker's acquisition and use of information it is necessary to understand the relationship between the information presented in texts and the representation of knowledge in human memory. Information is most readily assimilated when it is structured in a format that matches human cognitive structures and presented in a way that strengthens and maintains its memory representation. If a determination of human memory structures and processing strategies can be made, then it should be possible to adapt the format of presented texts to match these structures and hence facilitate the reader's assimilation and utilization of the information. In a real-world situation, this text formatting might be accomplished by filtering and reorganizing the to-be-learned material before it is presented to the decisionmaker, or by requesting the decisionmaker to process the information in the text in particular ways.

This report documents the results of a series of studies designed to elucidate the parameters of human learning from text. All the studies reported here consisted of controlled experiments in which subjects were presented with textual information, manipulated in a variety of ways. Subjects' performance on a number of tasks requiring the use of the textual information was measured.

These experiments served several useful purposes. First, they provided data on how various parameters of texts and learning environments influence the acquisition and effective use of information from the texts. This permitted the identification of text characteristics and learning situations that, when manipulated appropriately, can produce large fluctuations in a person's ability to learn and reason.

A second purpose of these experiments was to provide inferences about the knowledge structures people use to represent information from texts. The experiments supported the development of models of the underlying memory structures and processes required for task performance and allowed the evaluation of proposed models against alternative models. The development of the models was motivated by three specific goals: (1) to elucidate the process by which knowledge is acquired from a textual document, (2) to specify how the acquired knowledge is represented in memory, and (3) to identify how and when related information is integrated in memory. The theoretical determination of these underlying structures and processes was central to the research, because of the aforementioned working assumption that learning may be optimally facilitated by matching the structure of the input information to the preferred internal memory structures.

Finally, the results of these experiments were useful in suggesting and testing techniques for restructuring texts or the learning process in ways that improve such performance measures as reading time, amount of information learned, length of time the information is retained, and inferential ability.

Two broad task domains were chosen in which to study the structures and processes underlying human performance with texts. The first task domain was simple ACQUISITION, in which the reader attempts to learn and retain as much knowledge from the text as possible. The second task domain was INFERENTIAL REASONING, in which the reader must organize a set of facts from the text in order to generate or verify a conclusion following from the facts. For both task domains, the approach taken was to model the representation of knowledge in memory and the associated processes required to perform the task, and then to design text presentation formats that facilitated both the transfer of knowledge from texts into these memory structures and the performance of inferential reasoning.

Five sets of studies evaluated particular models and optimization techniques. Each study consisted of one or more experiments designed to determine the nature of the structures and processes underlying text learning and/or to evaluate the efficacy of various techniques for facilitating performance. The methodology, results, conclusions, and evaluation of proposed models of each of these studies is presented in detail in the following chapters. A brief description of each of these studies is given below.

In Study 1, a model of "schematic" learning was developed that provides an account of how knowledge in memory is used to guide the acquisition and organization of new, incoming information. The advantages of and constraints on the acquisition of new knowledge that conformed to a previously learned structure, or schema, were investigated experimentally. Subjects received various amounts of training on a set of structures by reading texts that utilized the structures; they then tried to learn a new set of texts that conformed to those structures. Memory for this new set was facilitated by prior experience with the text organizations, particularly when there was no confusion among the facts from different texts that shared a common organization.

In Study 2, information from newspaper stories was restructured into different formats in an attempt to affect memory for the stories. These formats included a condensed version of the news story (with all redundant and superfluous information deleted), a narrative structure based on the temporal order of events, a topical structure organized according to the primary concepts of the passage, and an outline structure that reduced the text to key phrases spatially arranged on the page in outline form. All forms of restructuring produced significant improvements in recall of the information contained in the passage.

Study 3 investigated the integration of related but separately acquired facts in memory either for apprehension of a complex idea or for performing inferential reasoning. Integration of separate facts into a unified memory structure was more likely to occur if related facts occurred in close temporal proximity and if the facts cued each other with similar wordings, rather than with paraphrased wordings. The integration of facts significantly improved the ability to recall related information and perform syllogistic reasoning. A model of knowledge representation and integration in memory was developed and evaluated against numerous alternative models.

Techniques for improving a learner's ability to organize diverse information for inferencing were investigated in Studies 4 and 5. In Study 4, subjects verified inferences based on information from two distinct texts. Performance was better when subjects had attempted to commit the texts to memory than when they were allowed to inspect the texts freely during the reasoning test. This result seems to indicate that memory can automatically organize related facts more reliably than conscious information-seeking searches of available, external sources. Based on these findings, a model of search and retrieval is proposed to account for these surprising results.

In Study 5, methods of annotating texts to facilitate integration and inferencing were investigated. Subjects read texts containing pairs of facts that could be integrated to support inferences. The second fact in each pair was annotated either with a footnote that repeated the first related fact, a footnote that contained the first fact and the appropriate inference following from the two facts, or no additional information. On a later test of reasoning, subjects who had received only the fact (but not the inference) annotation performed best.

The results of each study, are presented in the context of a set of models for knowledge representation and processing. Each study may stand alone in addressing a particular issue in knowledge representation. The underlying theoretical assumptions and conclusions following from each study are discussed separately in each chapter. However, taken together, these studies provide a unified corpus of research on related issues in learning and reasoning. The models, while addressing different task domains, share a common set of underlying assumptions about the structure of human memory. The individual studies are presented in detail in Chaps. II through VI. Based on these models and the obtained experimental results, a set of principles for improving human learning and performance with texts has emerged These principles are presented in Chap. VII, accompanied by brief descriptions of supporting data and references to the chapters in which particular results are discussed in detail.

II. THE USE OF SCHEMATA IN THE ACQUISITION AND TRANSFER OF KNOWLEDGE

Recent theoretical research on human memory has been stimulated by the rediscovery of the concept of the "memory schema." The notion of a schema was first introduced by Head (1920), who claimed that anything that enters consciousness is "charged with its relation to something that has gone before." Woodworth (1938), in his classic textbook on experimental psychology, extensively utilized the concept of a schema to describe various perceptual and memory phenomena. A memory schema, as it is typically conceptualized today, is a cluster of knowledge (a set of concepts and associations among the concepts) that defines a more complex and frequently encountered concept. A schema may represent anything from the componential features of a face (cf. Palmer, 1975) to the prototypical behaviors one engages in when going to a restaurant for a meal (Schank & Abelson, 1975). The concepts that constitute a schema may be perceptual features, semantic primitives, events or situations in the world, or, recursively, other schemata. Thus, schemata of varying levels of complexity coexist in memory.

The revival of interest in memory schemata as a theoretical construct is principally attributable to two lines of research. The first, conducted in the domain of artificial intelligence research, has sought to define new data structures for encoding complex descriptions of the world. The result has been a proliferation of knowledge representations that utilize some form of knowledge clustering such as "frames" (Minsky, 1975; Winograd, 1975; Kuipers, 1975), "scripts" (Schank & Abelson, 1975, 1977), or other forms of schemata (Schmidt, 1976; Moore & Newell, 1974). A second area in which schemata have received extensive treatment has been that of memory for connected discourse. Bartlett's (1932) seminal studies of prose memory led him to conclude that memory for a story consisted of a schema or plot framework and some associated details from the passage. Extending Bartlett's early ideas, several researchers have recently begun to use schemata to model the underlying memory structures that encode prose information (Rumelhart, 1975; Schank, 1975; Thorndyke, 1976, 1977, 1978; Kintsch, 1975; van Dijk & Kintsch, 1977; Winograd, 1977; Dooling & Christiaansen, 1977; Pichert & Anderson, 1977; Anderson, Spiro, & Anderson, 1977; Anderson, Reynolds, Schallert, & Goetz, 1976).

With this resurgence of interest in memory schemata, a few researchers have attempted to formulate general, yet comprehensive, memory models organized around the concept of knowledge schemata (e.g., Norman & Bobrow, 1975; Bobrow & Norman, 1975; Rumelhart & Ortony, 1977; R. C. Anderson, 1977). While each formulation of memory schemata has its unique characteristics, there seem to be a few properties common to each variant:

1. A schema represents a prototypical abstraction of the complex concept it represents. This abstraction contains a description of the composition and properties of the concept. For example, a "face" schema might specify that a typical face possesses two eyes, a nose, a mouth, and two ears, even though a particular face missing one or more of these features is still a face.

2. Schemata are induced from past experience with numerous exemplars of the complex concepts they represent. Presumably, we abstract the concept of a face after seeing many of them.

3. A schema can guide the organization of incoming information into clusters of knowledge that are "instantiations" of the schema. This represents a goal-directed focusing of processing by active memory schemata. Thus, when we catch a glimpse of a head, we activate our face schema and use the properties assumed by it to guide our search for features on the face we are viewing.

4. When one of the constituent concepts of a schema is missing in the input, its features can be inferred from "default values" in the schema. If the face is in shadow and we cannot see the mouth, we may still reasonably infer that it has two lips.

Previous attempts to formulate general "schema" theories of memory have had two principal shortcomings. First, the theories have been so vaguely specified or general that they are able to explain post hoc any set of available data. While many data may be taken to be consistent with schema theories, it is difficult to find any data that are inconsistent with them. Second, the theories proposed thus far have been used only descriptively to account for existing data. They have not been sufficiently well-specified to be used predictively. For example, it is not clear what a schema theory would predict about memory for an anomalous datum, i.e., a constituent detail in a set of information that did not fit the schema invoked to comprehend that information. Would it be well learned as a surprising stimulus (i.e., a von Restorf effect for schemata), or would it be poorly learned because it did not conform to the prototypical encoding structure?

The purpose of this chapter is to operationalize some of the theoretical constructs that underlie schema theory in a model of learning from simple texts. The model represents a marriage between previous work on the acquisition and sharing of memory schemata (Thorndyke, 1977) and the dynamics of learning and transfer of shared knowledge (Hayes-Roth, 1977). Many assumptions of the model draw upon classical research in learning. Some details of potential mechanisms for learning information in terms of schemata will be presented. The novel predictions of this learning model were tested in two prose learning experiments; the results are discussed later in this chapter.

In the remainder of this chapter, the term "schema" will be used to refer to a configuration of concepts and associations among the concepts that are repeatedly invoked to encode unique stimulus events. The concepts in the schema represent abstractions or generalizations of the originally presented concepts that invoke the schema. This configuration of concepts may represent knowledge exogenous to the syntactic and semantic relationships in the original inputs, as in the case of narrative schemata imposed by a reader on the events of a story to facilitate comprehension (Rumelhart, 1975; Thorndyke, 1977; Mandler & Johnson, 1977). In this case, the schema represents higher-order implicit relationships among concepts that would embellish a conventional propositional representation of the linguistic input. However, this study is concerned with a simpler form of schema, one in which the configuration of concepts represents a simple sharing of frequently repeated information. That is, the schema is an abstraction of a set of concepts and relationships that explicitly occur in a number of unique contexts. Because the schema is an abstraction of explicit knowledge, it has a structural integrity that is independent of any particular occurrence of the concepts that utilize it for their representation.

As an example, suppose one were to learn some information about the figures represented on Mount Rushmore. This information might be provided in a sequence of simple texts as follows:

Text 1: George Washington was the first President of the
United States. He lived in Mount Vernon.

Text 2: Thomas Jefferson was the third President of the
United States. He lived in Monticello.

Text 3: Abraham Lincoln was the 16th President of the
United States. He lived in a log cabin.

Text 4: Theodore Roosevelt was the 26th President of the
United States. He lived in Sagamore Hill.

A simplified representation of a portion of this information in a conventional memory model is given in Fig. 2.1. The occurrence of each text constitutes a new context, and none of the presented information stored in memory is shared between contexts. (Most propositional models would assume indirect associative connections between repeated occurrences of a concept such as "President" and "lived," mediated by direct associations through a common type node. However, only the individual occurrences, or "tokens," are depicted here, for simplicity.)

Much of the information in Texts 1 through 4 is identical. The predicates or relations in the texts are repeated, but the arguments to the predicates are different in each of the texts. That is, the knowledge in these texts might be conceptualized as "Text *i* states that person *i* was the $\underline{N}_i$-th President of the United States and

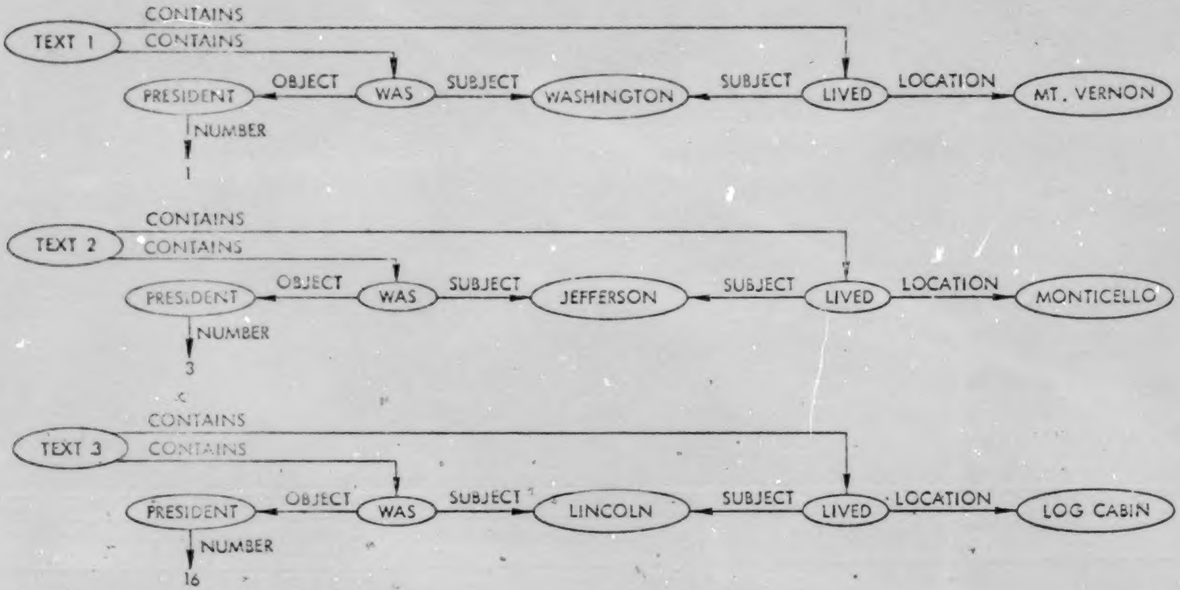
CONTAINS
TEXT 1
CONTAINS
PRESIDENT
OBJECT
WAS
SUBJECT
WASHINGTON
SUBJECT
LIVED
LOCATION
MT. VERNON
NUMBER
1
CONTAINS
TEXT 2
CONTAINS
PRESIDENT
OBJECT
WAS
SUBJECT
JEFFERSON
SUBJECT
LIVED
LOCATION
MONTICELLO
NUMBER
3
CONTAINS
TEXT 3
CONTAINS
PRESIDENT
OBJECT
WAS
SUBJECT
LINCOLN
SUBJECT
LIVED
LOCATION
LOG CABIN
NUMBER
16

Fig. 2.1--A hypothetical representation of the information in Texts 1-3 in a conventional propositional memory model

lived in location L_1. All texts share the common information but differ in the details of person, number, and location. If these texts were learned together, their respective representations might be integrated by the sharing of a common subrepresentation encoding the repeated information (see Chap. IV). This shared knowledge would constitute a schema in which the repeated concepts served as variables that could take on different values for each new occurrence. Figure 2.2 shows the resultant integrated structure and shared schema. The concepts labeled N_1, Person i, and L_1 represent placeholders, or slots, that are associated with (i.e., bind) different values for each new usage of the schema. In essence, these concepts are type nodes within the schema that subsume the tokens, or detailed values, that occurred in the various contexts. The context in which each detailed value occurred is preserved by means of a label on the link between the value and the abstracted concept that subsumes it. In Fig. 2.2

these links are labeled with the numbers of the texts in which the details occurred (1 through 4). These links may be thought of as context labels or time tags for the presented details, similar to those postulated in other propositional models (e.g., Anderson & Bower, 1973).

We assume that knowledge in memory is frequently organized with the use of such schemata. In particular, we believe that a knowledge structure formed in the representation of information from one context can be used to represent the same information occurring in different contexts (Hayes-Roth, 1977; Thorndyke, 1977). The use of a schema for encoding information depends, of course, on the successful activation of the relevant schema in memory at comprehension time. Such activation is probabilistic and depends upon such factors as the strength of the stored information, the extent of the overlap or match between input and schema, and the amount of time since the previous

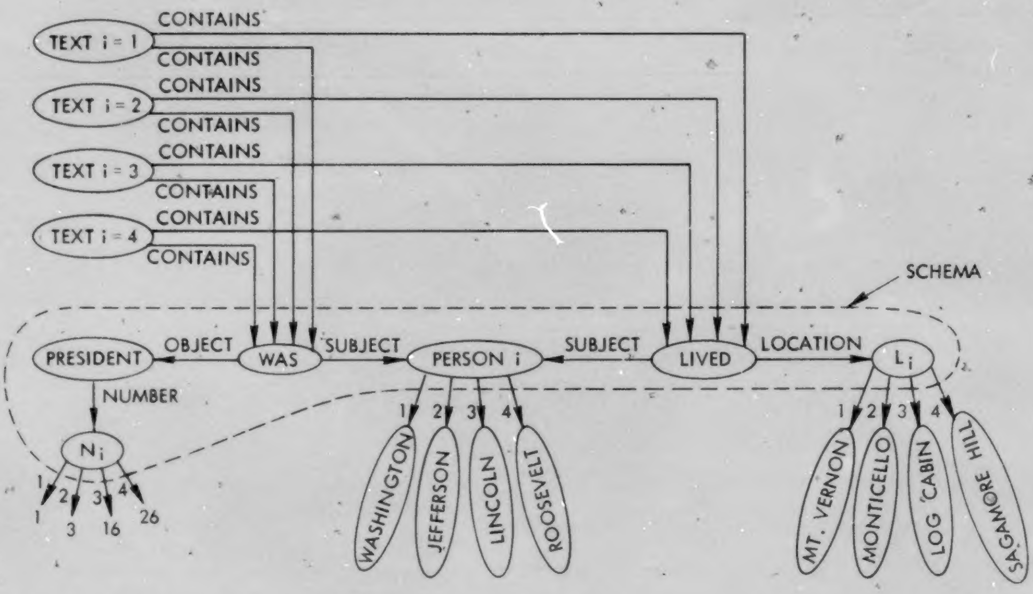
CONTAINS
TEXT i = 1
CONTAINS
CONTAINS
TEXT i = 2
CONTAINS
CONTAINS
TEXT i = 3
CONTAINS
CONTAINS
TEXT i = 4
CONTAINS
SCHEMA
PRESIDENT
OBJECT
WAS
SUBJECT
PERSON i
SUBJECT
LIVED
LOCATION
L i
NUMBER
N i
1 2 3 4
1 3 16 26
1 2 3 4
WASHINGTON
JEFFERSON
LINCOLN
ROOSEVELT
1 2 3 4
MT. VERNON
MONTICELLO
LOG CABIN
SAGAMORE HILL

Fig. 2.2--A shared schema for the representation of the common information in Texts 1-4 activation. These issues are treated elsewhere (Hayes-Roth, 1977, Chap. IV) and will not be addressed here. Rather, the concern here is to specify how the existence of such schemata influences the learning and retention of information that is represented with the use of the schemata. For example, given our assumption that the information in Texts 1 through 4 is represented as shown in Fig. 2.2, we wish to model how the acquisition of Text 4 will be influenced by the prior acquisition of Texts 1 through 3.

We assume that the use of a schema for encoding or retrieving information depends on its accessibility in memory (Tulving & Pearlstone, 1966; Hayes-Roth, 1977). The accessibility of a schema is the probability that it can be activated, either for use in storage of incoming information or for retrieval of previously stored information. Accessibility of a schema depends upon such factors as the strength of the stored information, the extent of the overlap or match between input and schema, and the recency and frequency of previous activations. Each time a schema is activated for use, it becomes more accessible for successive activations. The incremental effect of an activation on the accessibility of a schema is presumably a decreasing function of its prior strength. That is, a weak schema benefits more from an activation than a strong one. The assumption that accessibility of information increases with the frequency of activation and strength in memory has been proposed previously and has received considerable empirical support (Underwood, Runquist, & Schulz, 1959; Jung, 1963; Hayes-Roth & Hayes-Roth, 1975; Perlmutter, Sorce, & Myers, 1976; Hayes-Roth & Thorndyke, 1978).

When multiple details instantiate a variable concept in a schema, it is assumed that they compete with one another for associations with the variable concept. As the number of competing details increases, a person's ability to discriminate (and thus recall correctly) the context in which each detail occurred decreases. The discriminability among details associated with a single variable concept is assumed to be an inverse function of the number of competing concepts and the temporal proximity of their occurrences. The postulated effects of decreasing discriminability with increasing numbers of competing details are derived from previous work on interference in paired-associate transfer. In particular, these ideas resemble the notions of response competition (McGeogh, 1942; Osgood, 1949) and list-differentiation difficulty (Underwood, 1945; Underwood & Ekstrand, 1967; Anderson & Bower, 1972) in recall of paired associates. More recent studies have demonstrated that learning multiple propositions about a concept interferes with verifying any one of them (Thorndyke & Bower, 1974; Anderson, 1974; Hayes-Roth & Hayes-Roth, 1975).

Note that we postulate both benefits and costs associated with the use of schemata in learning. The benefits derive from the availability of previously learned representations of knowledge for use in the encoding of novel information. The strength of the prior information, the schema, incremented by the new activation required for the representation of the novel information should be greater than the strength of the representation of that information in a completely novel structure. On the other hand, the addition of novel information to the shared schema entails competition for associations among all the concepts necessarily sharing the same schema. Such competition should inhibit the acquisition process and produce interference in recall of the detailed concepts at retrieval time.

EXPERIMENT 1

These assumptions were tested in an experiment utilizing a transfer paradigm. Subjects learned a variable number of texts (the training texts) about a conceptual category, such as constellations. The topic of each text in the set was a different instance in that category, e.g., the constellations Pisces, Aries, Scorpio, etc. Subjects then learned and were tested on a transfer text (the target text) describing a new instance in the same category. Each sentence in the target passage had a corresponding sentence in all of the training passages that bore a particular relationship to it. For example, suppose the target passage contained the sentence "This constellation was originally charted at Palomar Observatory." The corresponding sentence in all previous training stories could then be one of three types: In the REPEATED condition, the entire sentence was repeated intact in all training stories (i.e., "This constellation was originally charted at Palomar Observatory"). In the CHANGED condition, the predicate was identical but a detailed case or argument to the predicate was changed in each of the training passages. For example, "This constellation was originally charted at Mount Wilson Observatory" might be one such prior sentence. These CHANGED sentences represented the same transfer situation as is shown in Fig. 2.2. In the UNRELATED condition, there was no similarity between the target sentence and any preceding training sentences. Thus in this condition, no sentences concerning the charting of the constellations would appear in any training story.

Subjects were tested for retention of the target story by either a free recall or cued recall test. On the cued recall test, subjects were given the predicate (e.g., "This constellation was charted at") and asked to recall the target detail. Using the assumptions given above, we may now predict performance on these tasks as a function of the number of training stories preceding the target story.

Figure 2.3a shows the qualitative effects on learning and retention of information from the target passage plotted against the number of training passages. The ordinate values above the origin indicate increasing positive effects on learning; the values below the origin indicate increasing negative effects on learning. During acquisition of the training passages, the repetition of substructures of the passages produces a representation in memory that is shared by all occurrences of that substructure. Repeated activation of the schema strengthens its representation, even when the details that instantiate it are unique in each context. Therefore, the accessibility of the schema for both storage and free recall of the target passage is a monotonically increasing function of the number of training passages. Because the benefit of an additional activation is a decreasing function of prior schema strength, the function is

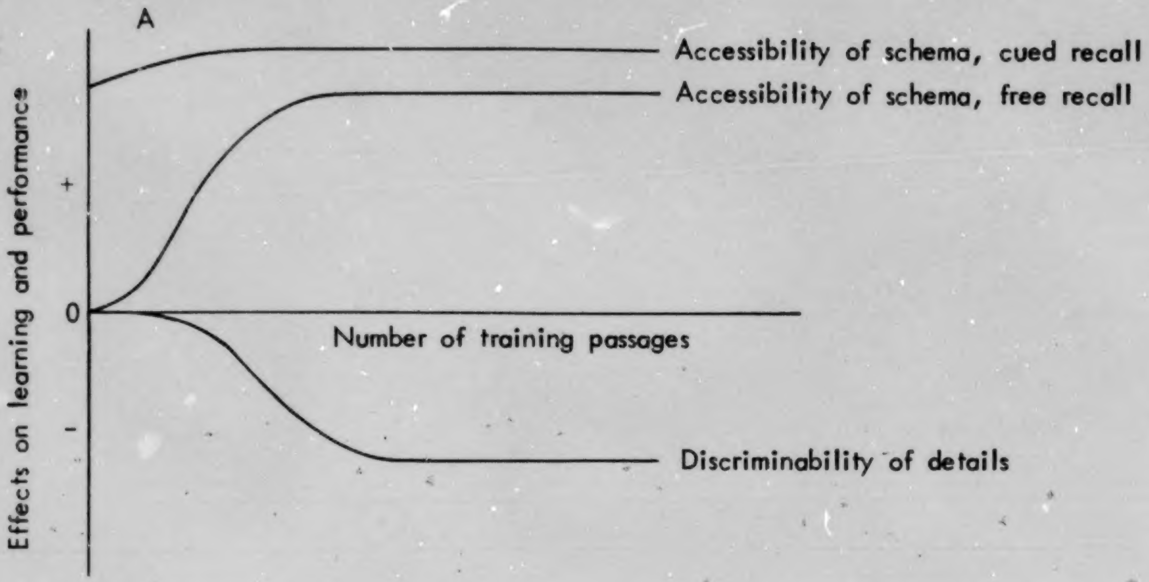

Fig. 2.3a--The theoretical functions for the effects on learning in Experiment 1

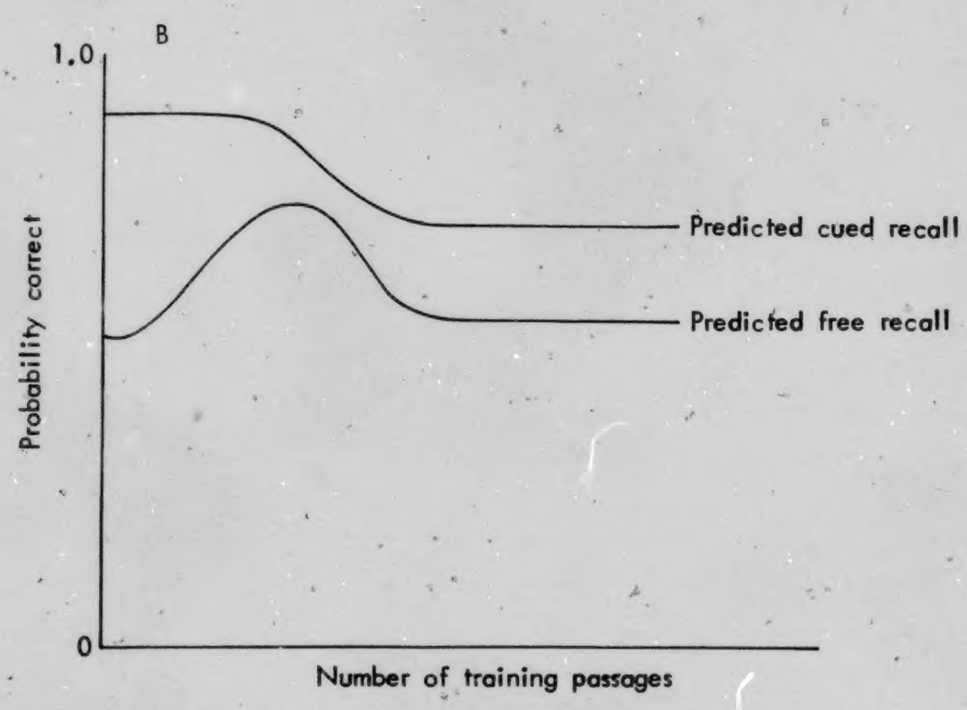

Fig. 2.3b--The predicted functions for free and cued recall of CHANGED sentences

negatively accelerated. That is, the slope of the function is steep in the early stages of learning and eventually reaches an asymptote when the schema becomes very well learned, as shown in Fig. 2.3a.

On the other hand, when details are changed in the successive occurrences of the schema (as in the CHANGED sentences), the discriminability of the context labels associated with the changed details produces a monotonically decreasing, or negative, effect on learning. This negative effect appears at some point following the establishment of the sharable schema representaion; that is, the schema must be established before it can be shared. Thus the positive effects of accessibility are initially stronger than the negative effects of decreased discriminability. The discriminability function presumably reaches asymptote at some point during training after the accessibility function reaches its asymptote.

The qualitative-effects functions shown in Fig. 2.3a may be operationalized as probabilistic behaviors defined over the range [0,1]. For example, the accessibility of the schema on the free recall task can be directly measured as the probability of recall of the sentence schema, or predicate. This probability may be designated as P(Predicate). Since the number of training passages increases from zero, this function begins at some value greater than zero and increases monotonically to an asymptote less than or equal to one. The detail discriminability function may be operationalized as the conditional probability of recalling the sentence detail, given recall of the predicate, or P(Detail/Predicate). This probability measures recall of the sentence detail, given that the predicate was successfully retrieved from memory. Differences in this probability across varying numbers of training passages are presumed to reflect only differences in a person's ability to retrieve the appropriate detail after the predicate was successfully retrieved. As the number of training passages increases, the function describing this probability is assumed to start at a value less than or equal to one and decrease to an asymptote greater than zero. Both functions are assumed to have non-negligible ranges. That is, there is measurable variation in both probabilities across training conditions. In addition, it is assumed that the domain over which the two functions vary is similar; that is, the variation between the times when the two functions reach their asymptotes is not vast.

When operationalized in this manner, these functions may be combined to predict subjects' performance on the recall tasks. Correct free recall of a sentence requires recall of both predicate and detail from the sentence. Thus the probability of sentence recall may be designated as P(Predicate and Detail). This probability may be expressed as the product of the two other probabilities:

$$P(\text{Predicate and Detail}) = P(\text{Predicate}) \times P(\text{Detail}/\text{Predicate}).$$

That is, the predicted function for CHANGED sentence recall is the product of the two effects functions in Fig. 2.3a. Given the assumptions listed above, the shape of this recall function should be that depicted by the lower line in Fig. 2.3b. As the number of training passages increases, recall of the CHANGED target sentences should initially increase, then decrease, eventually reaching asymptote. This result should reflect the increasing accessibility of the schema, coupled with increasing interference in recall of details. Since no assumptions were made about the starting or asymptotic values of the component probability functions, the absolute magnitudes of the component functions and the resulting free recall function cannot be predicted. Therefore, the relative values of the functions shown in Fig. 2.3b should not be taken literally. However, as long as the effects functions exhibit the depicted shape and satisfy the assumptions listed above, the qualitative shape of the free recall function may be predicted.

Similarly, we may predict the free recall function for the detail portion of the CHANGED sentences. This probability, designated as P(Detail), may be derived from the equation

$$P(\text{Predicate and Detail}) = P(\text{Detail}) \times P(\text{Predicate}/\text{Detail}).$$

The factor P(Predicate/Detail) is the probability of recall of the sentence predicate, given recall of the sentence detail. It is assumed that a subject rarely, if ever, recalls a detail without recalling the sentence schema in which it is embedded. Therefore, this probability should be one or very close to one and should not vary across different numbers of training passages. If this is the case, then the probability of recalling a CHANGED sentence detail, P(Detail), should be approximately equal to the probability of recalling the entire sentence, P(Predicate and Detail). Thus the functions describing these probabilities across varying numbers of training passages should be nearly identical.

For REPEATED items, the entire sentence was repeated intact in each of the training passages and the target passage. Discriminability of contexts would not be a problem in this case, since the information was identical in all contexts. Therefore, predicted free recall for REPEATED sentences should reflect a simple repetition effect--a monotonic increase with increasing numbers of training passages as predicted by the accessibility function alone. For UNRELATED items, no related sentences occurred in any of the training passages. Therefore, recall of these sentences should be lower than recall of CHANGED or REPEATED sentences and should not vary across increasing numbers of training passages.

On the cued recall test, subjects were given the sentence predicates as cues to recall the detailed case fillers. Each cue was the portion of the stored schema relevant for recall of the tested sentence. Thus the probability of accessibility of the schema at retrieval time would be essentially one and independent of the number of training passages. (Some small decrement in the accessibility of the cued schema might exist when small numbers of training passages had been presented, reflecting a weakly established memory representation.) This accessibility function for the cued recall test is shown as the top line in Fig. 2.3a. In this case, predicted performance for cued recall of CHANGED sentences is just P(Detail/Predicate), or the same as the discriminability function.

This predicted function is shown as the bottom line in Fig. 2.3b. Cued recall of CHANGED sentences should be a monotonic non-increasing function of the number of training passages. The cued recall function should initially be flat or slightly decreasing at a value greater than the peak of the free recall function. Then this function should exhibit a significant decrease, eventually ending at asymptote.

Method

Materials. Five to-be-recalled stories (hereafter referred to as "target" stories) were used. Each of these stories was unrelated to the others in topic and content. The titles of the five stories were "The Silicosis Disease," "The Apus Constellation," "The Circle Island Story," The John Payton Biography," and "The Filicules Plant." Each passage contained 12 sentences. For each target story, eight "training" passages were constructed that were different instantiations of the same general topic as the target passage. For example, the training passages for the Apus constellation story were about the Lepus constellation, the Pavo constellation, the Eranus constellation, and so on.

Each of the 12 sentences in a training passage corresponded to the sentence in the same serial position in its target story. The correspondence could be of one of three types: REPEATED, CHANGED, or UNRELATED. Each passage contained four sentences of each type. All sentences were approximately equivalent in semantic and syntactic complexity. For REPEATED sentences, the same fact was repeated verbatim in each of the eight training passages and target story. For example, if sentence 7 of the target passage was "It was originally charted at Palomar Observatory," then sentence 7 in each of the training passages would be identical. For CHANGED sentences, the same predicate was repeated in training and target passages, but in each instance it was instantiated with a different detail. So if sentence 7 in the target passage was in the CHANGED condition, one corresponding sentence in a training passage might be "It was originally charted at Mount Wilson Observatory." Thus in each of the eight training passages, the predicate "It was originally charted at ..." was instantiated with a unique detail. For UNRELATED sentences, there was no syntactic or semantic relationship among corresponding sentences in the training and target passages. Thus sentence 7 in two of the training passages might be "It was found to contain hydrogen gas" and "It is approximately 400 light years from earth."

For each subject, a new set of eight training passages for each target story was generated. First, the assignment of sentences in the target story to training condition (REPEATED, CHANGED, or UNRELATED) was randomized. Then, for each training passage to be constructed, the CHANGED and UNRELATED sentences were selected without replacement from the pool of 8 possible sentences in the designated serial position (1 through 12) and sentence condition (CHANGED or UNRELATED). Thus for each subject, the generation of new material provided a randomized assignment of items to condition and a randomized selection of training materials.

Subjects. One hundred UCLA undergraduates participated in the experiment, either for pay or to satisfy a course requirement.

Design. A 5 x 3 x 2 x 2 factorial design was used. Sentence type (REPEATED, CHANGED, or UNRELATED) and the number of training passages preceding a target passage were within-subject variables. Each of the five target stories was preceded by either 1, 2, 3, 4, or 8 training passages. The assignment of target story to training condition (1 through 8) was counterbalanced across subjects. Since each target story contained four sentences of each type, there were 20 of each sentence type per subject. One between-subject variable was the retention interval for the target passages (either 10 minutes or 24 hours). The other between-subject variable was test type (either free or cued recall). Different subjects performed these two tasks to insure against an artifactual effect of one task on performance of the other. The two groups defined by the two retention intervals will be referred to as the 0/0 group (0 hours between presentation of the training and target passages, 0-hour retention interval) and the 0/24 group (0-hour training-to-target interval, 24-hour target retention interval). Each of the 100 subjects was randomly assigned to one of the four groups so that there were 25 subjects in each group.

Procedure. Subjects were tested in groups of from one to eight persons. The experimental materials were included in three-part booklets provided to subjects, who worked at their own pace. Intentional learning instructions were given. Subjects were told to read the stories carefully because they would be tested on them later. Subjects then proceeded to Part I of the booklets, which contained the training passages. The stories were printed one per page. Subjects were allowed to read through the passages at their own pace but were not allowed to turn back to previous stories at any time. The first story was a buffer story, unrelated to all others in the experiment and identical for all subjects. The next 18 passages (1 + 2 + 3 + 4 + 8) were the training passages for the five target stories. Their order was randomized, with the constraint that one story of each type must occur in the last five serial positions. The final passage in the training sequence was unrelated to all others in the set and served as a buffer to minimize any recency effects in short-term memory. Following this final passage, subjects worked on multiplication problems that took approximately 10 minutes to complete.

In Part II, the intentional learning instructions were repeated. Then the five target stories were presented, one per page, in random order for each subject. The target stories were surrounded by unrelated buffer stories as in Part I. The study procedure was identical to that in Part I.

The 0/0 subjects then proceeded directly to Part III, the recall test for the target stories. The 0/24 subjects were dismissed until the next day, when they reconvened for the recall test. The target stories were tested in the same order as they had been presented in Part II. Subjects receiving the free recall test were instructed to write, for each story, as close to a verbatim recall of the story as they could. However, they were told not to omit anything that they remembered if they were unsure of exact wordings or sentence order.

Subjects receiving the cued recall test performed a sentence completion task for each target passage. The stories were presented as in Part II, with the detail instantiating each predicate omitted for each sentence. Subjects were instructed to fill in the missing portion with as close to the exact word or words as they could remember. Writing time for both free and cued recall tasks was unlimited.

Results

The results from the free and cued recall tests were analyzed separately. All data were initially analyzed using an analysis of variance that treated sentence type, number of training stories, and retention interval as factors. Arcsine transformations on proportions were used for the analysis to insure homogeneity of cell variances. For free recall, protocols were scored for gist recall of the presented information, with the exception noted below. For each sentence, it was determined whether the predicate had been recalled correctly, whether the detailed case filler had been recalled correctly, and whether the entire sentence (predicate plus detail) had been recalled correctly. In scoring recall of details, a paraphrased recall was counted correct only if it unambiguously specified the target detail and none of the corresponding training details. In cases in which paraphrase recall was impossible (e.g., recall of a year), exact recall was required for the trained item to be counted correct. A single scorer analyzed free recall protocols.

In each of the four groups (both free and cued recall in the 0/0 and 0/24 conditions), the mean number of UNRELATED items correctly recalled did not vary as a function of the number of prior training stories, as predicted by the theory above. As a result, for each subject, the mean number of UNRELATED items correctly recalled was averaged across the five training conditions (1, 2, 3, 4, 8) and was treated as a control or "0" training condition. That is, this value was used as an estimate of recall of a target item when no related sentences had preceded it during training.

The results for free recall of entire sentences are summarized in Fig. 2.4. Overall, subjects recalled more on the immediate test (0/0 group) than on the delayed test (0/24 group), F(1, 720) = 12.37, p < .001. The effect of number of training stories was significant, F(4, 720) = 14.78, p < .001, as was the type of sentence, F(2, 702) = 249.06, p < .001. For REPEATED sentences, recall increased with increasing number of training stories for both the 0/0 and 0/24 groups. As predicted, the recall of CHANGED sentences in the immediate test condition initially increases from the zero point, then

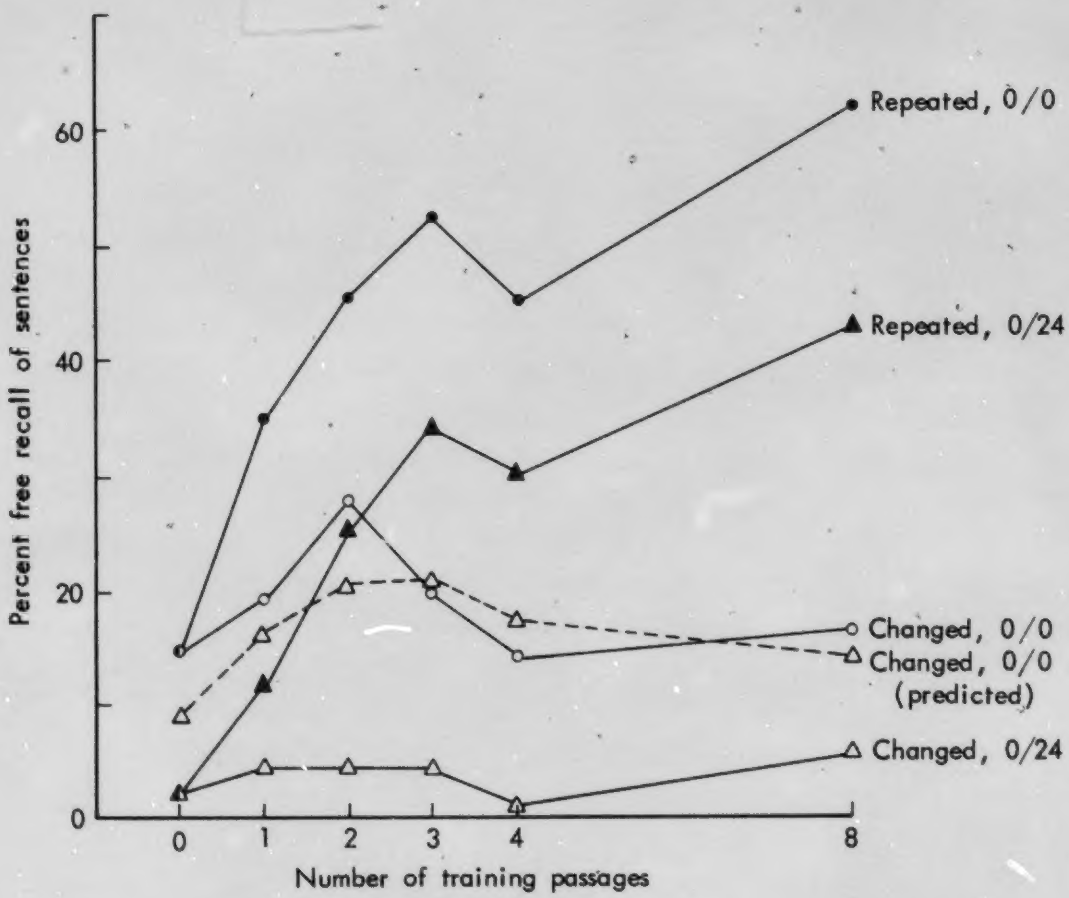

Fig. 2.4--Free recall of entire REPEATED and CHANGED sentences in Experiment 1

decreases, then reaches asymptote. Since no quantitative predictions were made regarding the exact location of the peak or the location of the asymptote relative to the zero point, there is no straightforward statistical test to evaluate the fit of the data to the predicted function. Several potential configurations of recall data would have been consistent with the prediction of an increase, followed by a decrease, followed by an asymptote. Therefore, the following Monte Carlo method was used to evaluate the reliability of this result.

For each subject, the mean recall scores of CHANGED sentences in the six training conditions were randomized and reassigned to the conditions. This was done for all 25 subjects, and the mean simulated recall curve was computed and plotted. This randomization was performed 100 times to produce 100 graphs of simulated recall. The graphs of the obtained data and the 100 randomizations were then rank-ordered for their fit to the predictions, by two independent judges, both of whom were thoroughly familiar with the theory and predictions but had not seen the obtained data. The rank order of the real data then constituted the probability that a fit to the predicted function by the data could be obtained by chance. For both raters, this probability was $p = .05$. While the immediate recall of CHANGED sentences showed the predicted inverted U-shaped function, after a retention interval of 24 hours the differences were eliminated and recall was very poor.

These results for free recall of sentences are broken down into separate recall of predicates and details in Fig. 2.5. For REPEATED sentences, immediate recall of predicates and details was virtually identical. In addition, free recall of predicates from CHANGED sentences was as accurate as recall of REPEATED predicates, $t(4) = 1.82$, n.s. After a 24-hour retention interval, recall of predicates dropped significantly across all training conditions but still reflected the increasing effect on recall of number of training stories. This is shown by the line for recall of the CHANGED predicates in the 0/24 condition in Fig. 2.5.

While recall of predicates for the CHANGED sentences increased with number of training stories in both immediate and delay conditions, immediate recall of CHANGED details increased, then decreased to asymptote. Thus the sentence free recall function in Fig. 2.4 for CHANGED items reflected fluctuations in recall of details, not predicates. The fit of the recall function for CHANGED details in Fig. 2.5 to the predicted function was tested using the same randomization method for details as was described above for sentence recall. The attained significance levels for the recall data thus obtained from the two independent raters were p = .04 and p =.05.

The intrusions of CHANGED details from training stories into recall of CHANGED target passages are also shown in Fig. 2.5. While

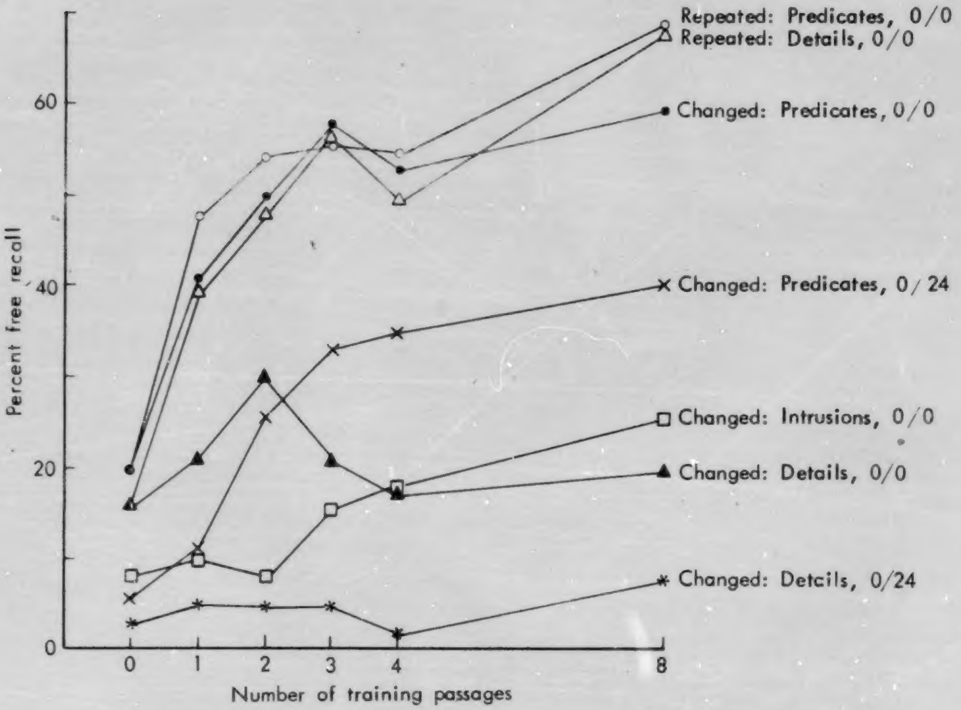

Fig. 2.5--Free recall of sentence constituents (predicates and details) in Experiment 1

correct recall of details increased, intrusions did not vary from their base, or zero, level. (The zero point is the probability of intruding an UNRELATED sentence from a training passage into recall of the target passage.) When correct recall of details decreased, intrusions increased reliably, $F(5, 144) = 2.31$, $p < .05$. Combining these two functions provides a measure of the probability of a schema-relevant response for CHANGED sentences as a function of the number of preceding training passages. That is, by adding the intrusion and correct recall probabilities for details, we obtain the probability of recalling any detail that was associated with a variable concept during training. This function increases reliably from 24 percent to 45 percent, $F(5, 144) = 2.46$, $p < .05$.

The results for cued recall are summarized in Fig. 2.6. As in free recall, performance is better on the immediate than on the delayed test, $F(1, 720) = 12.67$, $p < .001$. Reliable differences were obtained due to both sentence type, $F(2, 720) = 214.76$, $p < .001$, and training condition, $F(4, 720) = 10.56$, $p < .001$. As predicted, the cued recall of CHANGED details in the immediate test condition is initially flat and then decreases as the number of training stories increases. This decrease in cued recall is reliable, $F(5, 144) = 2.61$, $p < .05$. By comparing Figs. 2.5 and 2.6, the relative levels of free and cued recall of CHANGED details may be noted. As predicted, the initial flat portion of the cued recall curve (40 percent) is higher than the highest point on the free recall curve (30 percent). For delayed cued recall, as for free recall, performance on CHANGED details was poor and did not vary across training conditions.

As shown in Fig. 2.6, intrusions in cued recall of CHANGED details increased monotonically with number of prior training stories. While this result was in the expected direction, it failed to achieve significance. As for free recall, the probability of generating a schema-relevant response to CHANGED predicate cues was computed by adding the correct recall probability and intrusion probability in each training condition. There was no difference in the resulting

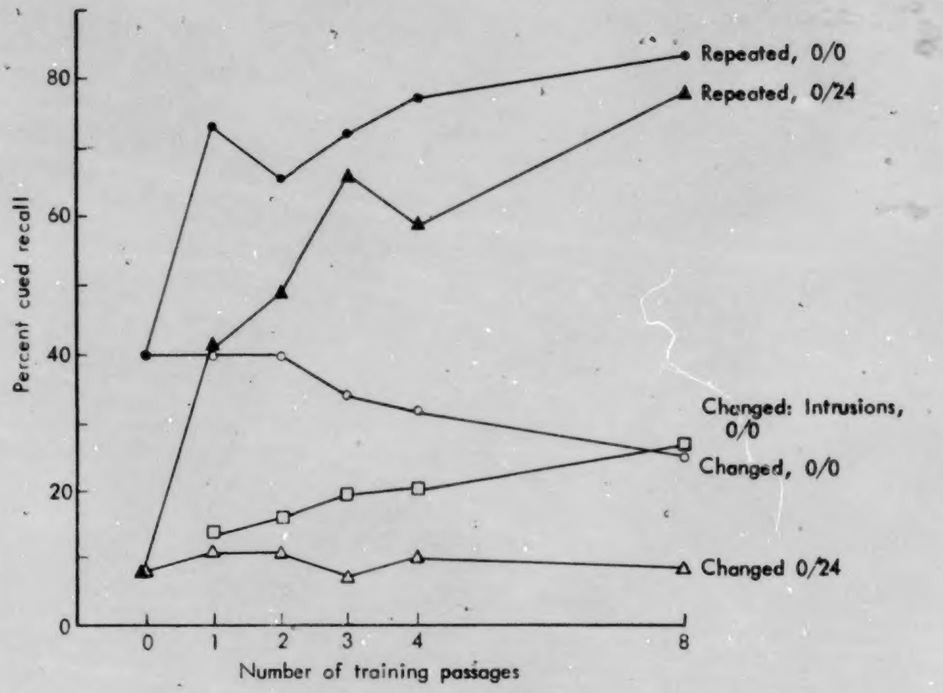

Fig. 2.6--Cued recall of details for REPEATED and CHANGED sentences in Experiment 1

response probability as a function of number of training stories; the probabilities for the five training conditions were all within the range of 50 to 56 percent.

Discussion

These results confirm the predictions discussed above. During the acquisition phase of the experiment, subjects constructed schemata representing common information in related passages. With each new training passage, the schema representing the shared information would be activated and strengthened. For REPEATED sentences, this activation produced a strengthening of the entire repesentation of that sentence, as would be predicted by many learning theories. For CHANGED sentences, the common predicate would be strengthened, but the detail instantiating the predicate would compete for association with that predicate with other details from different contexts. The accessibility of the shared information was measured directly from free recall of the repeated information. As expected, accessibility increased monotonically with increased repetitions of the shared schema. The incremental accessibility of repeated information was independent of whether that information had multiple associates (CHANGED predicates) or a single repeated associate (REPEATED predicates).

Immediate free recall of entire CHANGED sentences and CHANGED details confirmed the novel predicted inverted U-shaped function. This function derived from the combined effects of increased accessibility of the shared information and decreased discriminability of the details as the number of training passages increased. The claim that this function represents the product of the accessibility and discriminability functions may be evaluated by combining the data from the two functions. The accessibility function was operationalized as free recall of CHANGED sentence predicates. As expected, recall increased monotonically with increased repetitions of the shared schema. The incremental accessibility of repeated information was independent of whether that information had multiple associates (CHANGED predicates) or a single repeated associate (REPEATED predicates).

The discriminibility effect function can be estimated from the cued recall data. Since providing the cue at test time equalized or nearly equalized accessibility to the schema for the various training conditions, recall of CHANGED details would presumably reflect only the conditional probability P(Detail/Predicate). This was the operationalization of the discriminability-effect function shown in Fig. 2.3a. This function predicts little change in discriminability as the shared schema becomes established in memory, followed by steadily decreasing discriminability among the details as more and more of them become associated with the shared predicates. As shown in Fig. 2.6, the cued recall of CHANGED details shows this exact trend. These two component functions, the free recall of CHANGED predicates and cued recall of CHANGED details, were obtained from different subject samples. The product of these two functions estimates the predicted function for free recall of CHANGED sentences. This function is shown by the dashed lines in Fig. 2.4. The function is nearly identical to the obtained free recall function and did not differ reliably from it, $\underline{t}(5) = .58$, n.s.

The intrusion data in free recall may also be used to estimate the discriminability effect function. Presumably, an intrusion occurs when the subject successfully activates the schema for recall but cannot discriminate the target detail from others he has studied, thereby retrieving an incorrect detail. According to the predicted-effect function in Fig. 2.3a, the difficulties in discriminability should be negligible for small numbers of training passages and then steadily increase with greater numbers of training passages. That is, as the number of training passages increases, intrusions should initially stay the same, then increase. This precise result was obtained in free recall, as shown in Fig. 2.5.

According to the model of schemata proposed above, the accessibility of a schema should decrease as a function of the time since its previous activation. Therefore, in general, the longer the retention interval, the lower the accessibility of the schema in memory. In addition, the longer the retention interval, the more pronounced should be the negative effect of discriminability of details sharing the schema. As the retention interval of shared knowledge increases, then, the sum of these two effects functions should flatten and depress the recall function. This effect was obtained in free recall of CHANGED sentences in the 0/24 condition, where recall was worse than in the immediate test condition and did not vary over training conditions.

If the sharing of details from different contexts by a schema exerts a negative effect on learning and retention, then increasing the discriminability between the target and training contexts should improve performance on recall of the target facts. One technique for improving this discriminability would be to decrease the temporal proximity of the training and target passages. If during learning of the CHANGED sentences there were no negative effects of decreased discriminability with increasing numbers of training passages, then no interference in learning of CHANGED details should be obtained by increasing the number of training passages. This hypothesis was tested in Experiment 2.

EXPERIMENT 2

The materials and methodology of Experiment 2 were identical to those of Experiment 1, with one important exception. After presentation of the training materials, subjects waited 24 hours before receiving the target passages. This 24-hour delay presumably would increase the discriminability between the training and target details of CHANGED sentences sharing single predicates in the schema, relative to the case in which target presentation immediately followed training. Of course, even the discriminability produced by the 24-hour interval would fade over a long retention interval. On an immediate test, however, the theory would predict that the discriminability-effect function in Fig. 2.3a should be flat at zero across the various training conditions.

We now consider the predicted-accessibility function in this situation. At the time of target passage presentation, 24 hours would have elapsed since the previous activation of the experimental schemata. The shape of the accessibility function after 24 hours may be estimated from the graph of REPEATED sentence recall in the 0/24 condition of Experiment 1 (Fig. 2.4) and the graph of the CHANGED predicate recall in the 0/24 condition (Fig. 2.5). After this retention period, the accessibility function still increased monotonically with increasing training passages, even though the strengths of the schemata had faded over the retention period. Presentation of the target passages would activate and thus strengthen the schemata. However, the increment in accessibility would be a decreasing function of prior strength. Therefore, the accessibility function, while remaining monotonically increasing across training conditions, should be flattened considerably.

Free recall of entire CHANGED sentences and CHANGED details should be the product of the accessibility- and discriminability-effects functions. In particular, as the number of training passages increases, free recall of CHANGED sentences and details should increase monotonically and should be proportional to recall of CHANGED predicates. Since the CHANGED predicates would have been presented n times and the target CHANGED details would have been presented only once, the former would have greater strength in memory and therefore a higher overall probability of being recalled. However, over increasing numbers of training passages, recall of CHANGED details should increase at the same rate as recall of CHANGED predicates.

In cued recall, the effect of the cue should be to provide accessibility to the stored schema. Any differential effects of accessibility due to training conditions should be removed, and thus cued recall of CHANGED details should reflect only the discriminability-effect function. This effect is predicted to be negligible across training conditions, so cued recall of CHANGED details should not vary. Another estimate of the discriminability function may be obtained from the intrusion errors for CHANGED items on both free and cued recall. For each of these measures, there should be no differences due to training condition.

Method

Materials. The materials were identical to those used in Experiment 1. Each of the five target passages contained 12 sentences, 4 each that were REPEATED, CHANGED, or UNRELATED with respect to the training passages that preceded them.

Subjects. One hundred UCLA undergraduates participated in the experiment, either for pay or to satisfy a course requirement.

Design. The same design as in Experiment 1 was used. The dependent variables were percent free recall and percent cued recall of sentences from the target passages. The number of training passages preceding a target passage (1, 2, 3, 4, or 8) and sentence type (REPEATED, CHANGED, or UNRELATED) were within-subject variables. Test type (free or cued recall) and retention interval for the target passages (0 or 24 hours) were between-subject variables. The two groups defined by the two retention intervals will be referred to as the 24/0 group (24 hours between presentation of the training and target passages, 0-hour retention interval) and the 24/24 group (24-hour training-to-target interval, 24-hour target retention interval). Each of the 100 subjects was randomly assigned to one of the four groups so that there were 25 subjects in each group.

Procedure. The study and test procedure was the same as in Experiment 1. After studying to learn the training stories presented in booklets (Part I), subjects were dismissed and asked to return at the same time the next day. In the second session (Part II), the intentional learning instructions were repeated. Then subjects read the target stories as in Experiment 1. The 24/0 subjects then proceeded directly to Part III, the recall test for the target stories. The 24/24 subjects were dismissed until the next day, when they reconvened for the recall test. The target stories were tested in the same manner as in Experiment 1.

Results and Discussion

The cued and free recall results were initially analyzed separately using an analysis of variance. Free recall protocols were scored as in Experiment 1. In each of the four groups (both free and cued recall in the 24/0 and 24/24 conditions), the mean number of UNRELATED items correctly recalled did not vary as a function of the number of prior training stories. Therefore, for each subject, mean performance on UNRELATED predicates and details in both retention conditions was averaged across the five training conditions (1, 2, 3, 4, 8) and treated as a zero training condition, as in Experiment 1.

The results for free recall of entire sentences are summarized in Fig. 2.7. Subjects recalled more on the immediate test (24/0 group) than on the delayed test (24/24 group), $F(1, 720) = 10.12$, $p < .001$. The effect of number of training stories was also significant, $F(4, 720) = 11.43$, $p < .001$, as was the type of sentence, $F(2, 720) = 187.58$, $p < .001$. Recall of REPEATED sentences increased with increasing numbers of training stories on both immediate and delayed tests. As predicted, immediate recall of CHANGED sentences (the 24/0 condition) also increased monotonically with increased number of training stories. Although this function was in the predicted direction, a planned comparison failed to confirm a linear trend in the data, $F(1\ 144) = 1.99$, n.s. Recall of CHANGED sentences on the delayed test (24/24 condition) was depressed and constant across training conditions.

One method for assessing the effect of the delay between training and target passage presentation on improving discriminability is to compare the CHANGED recall results from Experiments 1 and 2. Overall, free recall of CHANGED sentences in the 24/0 condition was superior to recall in the 0/0 condition (no training-to-target delay), $t(4) = 6.72$, $p < .01$. Moreover, this superiority in recall was maintained after a 24-hour retention interval. That is, recall of CHANGED sentences in the 24/24 condition was reliably better than in the 0/24 condition, $t(4) = 2.39$, $p < .05$.

To determine the effects on learning and recall of the CHANGED sentences in the 24/0 condition of Experiment 2, these recall data were broken down into predicate and detail recall graphs. These results are shown in Fig. 2.8. CHANGED predicates were recalled reliably better than details, $F(1,288) = 9.69$, $p < .001$. This result was expected, since predicates presumably received more frequent activations, and hence were more accessible, than target details. For both predicate and detail recall, performance increased with increasing numbers of training stories. This increase was

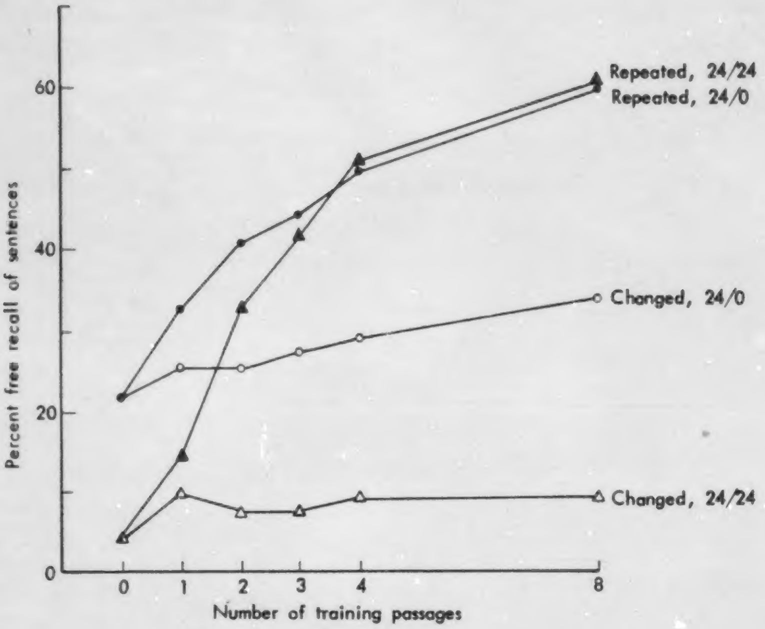

Fig. 2.7--Free recall of REPEATED and CHANGED sentences in Experiment 2

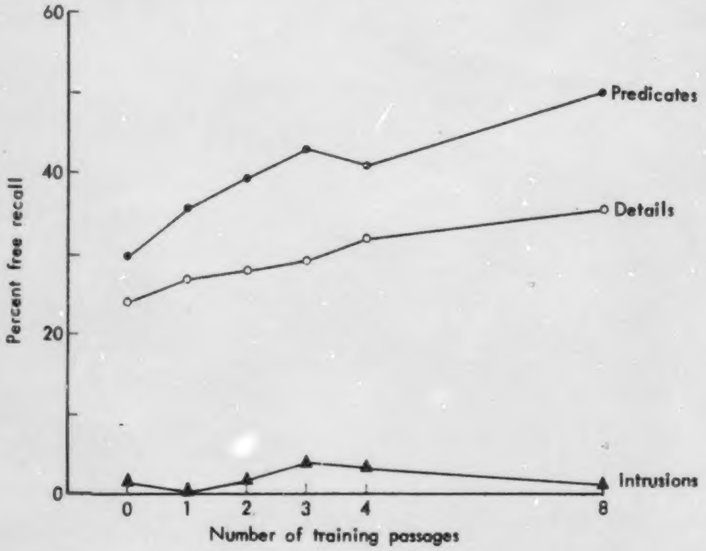

Fig. 2.8--Immediate free recall of the CHANGED sentence constituents(24/0 condition) in Experiment 2

reliable, $\underline{F}(1,288) = 4.00$, $\underline{p} < .05$. Furthermore, the interaction between item type (predicate or detail) and training condition was not significant, $\underline{F}(5, 288) = 1.26$, n.s. This failure to find a significant interaction is an important result for two reasons. First, it indicates that the significant increase in recall across training conditions is attributable to increases in detail recall as well as predicate recall. Second, this result implies that detail recall is proportional to predicate recall, as predicted. Thus the increase in detail recall can be attributed to the increase in schema accessibility and the absence of negative effects of discriminability.

Additional support for this latter conclusion may be adduced from the intrusion data for CHANGED sentences. It may be noted in Fig. 2.8 that subjects virtually never intruded a detail learned during training into recall of target CHANGED sentences. Furthermore, there was no increase in the intrusion rate as the number of training passages increased. Thus the 24-hour delay between training and target presentations guaranteed the elimination of discriminability difficulties.

The cued recall results for Experiment 2 are shown in Fig. 2.9. As in free recall, performance was better on the immediate than on the delayed test, $\underline{F}(1, 720) = 14.13$, $\underline{p} < .001$. Cued recall of details from REPEATED sentences increased with increasing repetitions provided by the training passages. This result was obtained in both the immediate (24/0) and delayed (24/24) test conditions. As predicted, cued recall of CHANGED details in the immediate test condition did not vary significantly across training conditions. Cued recall level was constant for 0 to 4 training stories. Performance increased for the 8 condition, but a post-hoc Newman-Keuls test declared this difference to be unreliable ($\underline{p} > .10$). To reject the hypothesis that the 24-hour training-to-target interval did not alter detail discriminability, the cued recall results for CHANGED sentences in the 0/0 and 24/0 conditions were compared. Overall, cued recall was better in the 24/0 than in the 0/0 condition, $\underline{F}(1, 47) = 5.30$, $\underline{p} <.05$.

In addition, the decrease in cued recall across training conditions in the 0/0 condition produced a significant interaction, F(4, 188) = 2.50, $p < .05$.

Intrusions of training details in cued recall of CHANGED sentences on the immediate test did not vary significantly across training conditions. This result was as predicted and provides additional evidence for the assumption that the training-to-target delay improved the discriminability between the training and target details.

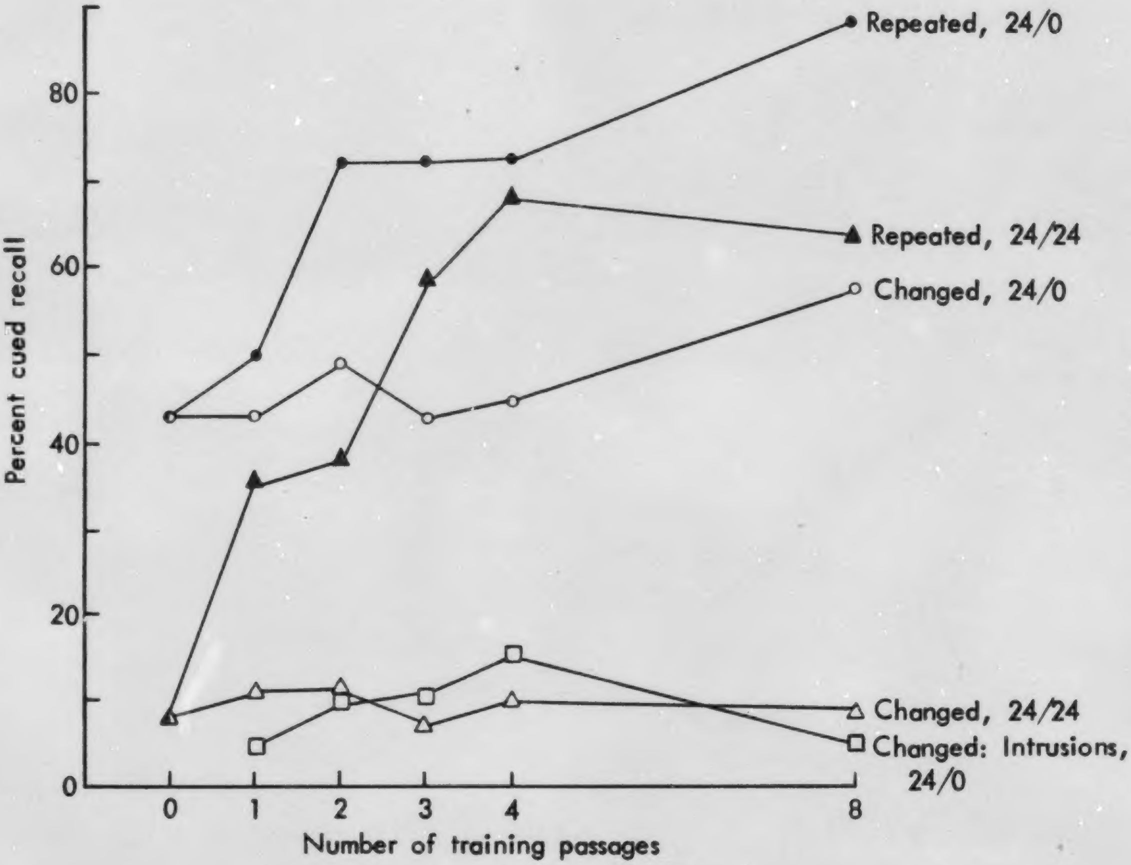

Fig. 2.9--Cued recall of details for REPEATED and CHANGED sentences in Experiment 2

GENERAL DISCUSSION

These results provide consistent support for the proposed model of memory schemata. That model presumes that knowledge substructures are shared in memory to encode information acquired in different contexts. The model rests on some basic assumptions about the structure of these schemata in memory and the processes that operate on them. The major assumptions and the results that bear on those assumptions are briefly summarized below.

1. The use of a schema or shared substructure in a number of unique contexts strengthens the represenation of the shared information and increases its accessibility in memory. This prediction is common to many previous theories of learning (e.g., Hebb, 1949; Anderson & Bower, 1973). For both REPEATED and CHANGED sentences and in all retention conditions (0/0, 0/24, 24/0, 24/24), free recall of the shared information (the sentence predicates) increased with increasing numbers of training passages. These training passages constituted repetitions of common information in new contexts with either a single or varied associates.

2. When information from different contexts shares the same schema in memory, there is interference in learning and retention of the subset of the information that is unique to each of the different contexts. This interference increases with increasing numbers of competing associates to the schema. This prediction was confirmed by the data in the 0/0 condition for intrusions in both free and cued recall of CHANGED details and by the data for correct cued recall of CHANGED details. The probability of correct cued recall of details decreased with increasing numbers of other details competing for associations with the same schema. Intrusions of those competitors, on the other hand, increased across training conditions.

3. The facilitative and inhibitory effects of the use of schemata for learning combine to predict acquisition and retention. As a shared schema is strengthened through its use in multiple unique contexts, acquisition of information sharing that schema is initially facilitated, then inhibited, and finally unaffected as the

componential effects reach their maximal values. In free recall of CHANGED sentences and CHANGED details in the 0/0 condition, this prediction (shown in Fig. 2.3) was confirmed. As the number of prior training passages increased, recall of CHANGED items first increased, then decreased to asymptote.

4. The interference among competing associates to a schema can be reduced or eliminated by increasing the discriminability among the various contexts of occurrence of the competing information. When the presentation of the target passages followed the training passages by 24 hours, immediate recall of CHANGED sentences and details (the 24/0 condition) was improved relative to the no-delay condition (0/0). In addition, the reduction of discriminability difficulties did not come at the expense of the facilitative effect of schema repetition: Recall of CHANGED items increased with increasing numbers of training passages.

Our characterization of memory schemata addresses but a single level of complexity in what is shared and transferred. A schema in this study was operationalized as a semantic predicate or relation and the associated case frames that instantiated the predicate. The use of a schema in multiple contexts consisted of the repetition of sets of these predicates in different passages with either the same (REPEATED) or different (CHANGED) case fillers. Other researchers have studied the acquisition and transfer of both more and less complex schemata.

Hayes-Roth (1977) predicted and obtained the combined transfer function in Fig. 2.3 using paired-associate nouns as materials. Subjects were given variable amounts of training on A-B pairs and were then transferred to learning of A-C pairs. A recognition test measured recognition confidence on A-C pairs as a function of A-B learning. At low levels of A-B learning, establishment of the A-C representation benefited from the accessibility of A in memory. However, with increased strength of the A-B association, activation of A for use in A-C learning entailed the simultaneous activation of B, thus interfering with establishment of the A-C representation. In that study, the shared information was simply the noun A, but transfer effects similar to the present ones were obtained.

Thorndyke (1977) investigated the transfer of a much more abstract learning schema in a study of text memory. Subjects were presented a narrative text to learn for a later recall test. According to a proposed theory of text learning, the comprehension and assimilation of the text into memory required the use of a schema for story structures that facilitated the organization of simple propositions from the story into higher-order functional units. These units reflected the integration of the story information into a coherent knowledge representation of the plot and episode structure of the narrative sequence as well as the semantic content of individual propositions. The to-be-learned story was preceded either by a story with unrelated content but an identical narrative schema, or by a story unrelated in structure and content. It was postulated that in the repeated structure condition, the story schema encoded during first-story learning could be successfully utilized for representing the to-be-learned story. In fact, this repetition produced a significant 22 percent improvement in learning, compared to the control condition. Thus, with one prior learning trial on a story schema, positive transfer of the schema to a new context was obtained. Similar effects of proactive facilitation have been found using expository educational materials (Royer & Cable, 1975, 1976).

The observation that multiple associations to a knowledge unit produce interference is, of course, a well-established result. The hypothesis that multiple associates to a knowledge unit produce interference because of discriminability difficulties was first proposed by Underwood (1945). Recently, numerous researchers (Crouse, 1971; Anderson & Myrow, 1971; Bower, 1974; Kuhara, 1976) have found retroactive interference in the recall of detailed facts from prose passages when the interpolated passages contain the same facts with new details (as in the CHANGED sentences used here). In addition, Bower (1974) found retroactive facilitation in recall of information repeated in the interpolated passages either with the same or with new associates (as in the REPEATED and CHANGED predicates used here). However, none of these researchers predicted the combined facilitation-interference function for performance that was obtained here for CHANGED sentences.

One potentially important factor that was not manipulated in the present experiment is degree of learning of each individual CHANGED sentence. Hayes-Roth (1977) has argued that the qualitative nature of transfer effects (positive, negative, or null) is largely determined by the degree of learning of the training material. In particular, she demonstrated that (1) minimally learned training material produces primarily positive transfer effects; (2) moderately well-learned training material produces primarily negative transfer effects; and (3) overlearned training material produces no transfer effects. These effects presumably reflect changes in the availability of memory substructures for use in encoding new information. In the present experiments, all CHANGED training sentences were presented only once and were thus minimally learned. If the training stories had been better learned, the observed transfer effects might have been qualitatively different. Whether or not an interaction would occur in the present paradigm remains an empirical question.

Most of the attempts made to date to develop schema theory as a viable psychological theory have focused on the representation of well-learned schemata in memory (Rumelhart, 1975; Schank, 1975; Thorndyke, 1977, 1978; Rumelhart & Ortony, 1977) or the processes that operate in conjunction with schemata during comprehension or memory search (Norman & Bobrow, 1975; Bobrow & Norman, 1975; Rumelhart & Ortony, 1977). To date, schema theorists have said little about the dynamics of and constraints on the acquisition of new information *per se*. The tacit assumption of the class of schema theories seems to be that new information is acquired by producing a copy (or token) of an existing schema in memory and "interpreting" the new information in terms of the schema by instantiating as many of the variable concepts as possible (cf. Rumelhart & Ortony, 1977). While this is a reasonable theoretical generalization upon our learning theory, it is as yet empirically untested.

In contrast, Experiments 1 and 2 focused primarily on those aspects of schemata that are relevant for the learning process. In essence, we have proposed a learning theory that combines the new notions of memory schemata with some traditional psychological assumptions about learning. Our theory of schemata as shared knowledge structures is similar to other formulations of schema theory; however, it goes beyond them as a psychological theory by detailing the costs and benefits associated with the use of schemata in learning.

III. KNOWLEDGE ACQUISITION FROM NEWSPAPER STORIES

Chapter II investigated the effects on learning of repeating shared knowledge substructures. This chapter investigates the relative efficacy of particular text organizations for learning. The influence of text organizations on the acquisition of information has received extensive treatment by educational and cognitive psychologists. In some studies, the investigators identify the learning objectives for subjects--that is, the subset of text information to be sought and learned (e.g., Gagne & Rothkopf, 1975). More typically, however, a reader attempts to learn all the information in the passage, since he does not know what particular facts will be of use later. In such cases, the structure of the material is a critical determinant of what will be learned from the passage. In general, familiarity with the structure of the material, independent of its semantic content, can facilitate learning of the material (Thorndyke, 1977).

Several studies have investigated the efficacy of various information organizations for text learning. In a series of related studies (Frase, 1969b, 1973; Friedman & Greitzer, 1972; Myers, Pezdek, & Coulson, 1973; Perlmutter & Royer, 1973; Schultz & DiVesta, 1972), text passages were constructed from matrices of name-attribute-value triples. That is, a set of concepts in a semantic category were described using a fixed set of attributes, and each concept in the category had different values of the attributes. In general, these studies demonstrated that any coherent organization of the information (i.e., either by names or by attributes) produced better learning than a random organization, and that the two organizations produced equal learning.

The texts used in these studies described a small, fixed set of exemplars from a single conceptual category, with each exemplar possessing different values for a particular set of attributes. More typically, a text that someone has to learn contains information about multiple topics, has a wide variety of predications about those topics, and contains events or actions with a temporal structure. Traditional classification systems segregate prose passages into four basic genres: description, exposition, narrative, and persuasion (Brooks & Warren, 1972). Depending on the author's intention and point of view, texts of the first three types are frequently organized either topically, with information organized around conceptual themes, or temporally, with information presented in a narrative sequence. These two organizations might have very different consequences for how well a passage can be learned, particularly if one of the organizations is preferred by readers.

In studies of the effects of these two forms of organization on learning (Sasson, 1971; Kulhavy, Schmid, & Walker, 1977), conceptual organization of information was superior to temporal organization. In the Sasson study, however, the texts were presented to subjects one word at a time in serial order. This procedure bears little resemblance to the way in which people normally study and learn texts, so conclusions from this study regarding normal processing modes must be regarded as tenuous. In the Kulhavy et al. study, only recall of a few target words was measured, so it is difficult to assess the effect of the organization manipulation on overall learning of a coherent prose passage.

In attempts to precisely model the organization of textual information in memory, several researchers have proposed detailed representations for the structure in a prose passage (Kintsch, 1974; Meyer, 1975; Rumelhart, 1975; Frederiksen, 1975; Thorndyke, 1977; Mandler & Johnson, 1977). Various predictions from these models for the influence of structure on acquisition and retention of information have been tested empirically. A typical finding is that the "centrality" or "importance" of a proposition to the "theme" of the passage predicts the recall probability of the proposition (Kintsch, 1974; Meyer, 1975; Thorndyke, 1977). From these and other studies, a general theory of "schemata" has emerged (Anderson, 1977; Rumelhart &

Ortony, 1977; Thorndyke, 1978). This theory assumes that a person has in memory a set of prototypical structures for use in comprehending and encoding prose information, particularly goal-directed narrative sequences. For narrative stories, these structures organize a temporal sequence of actions into a hierarchically arranged set of episodes that reflect their causes and consequences (Rumelhart, 1975; Mandler & Johnson, 1977; Thorndyke, 1977). A common assumption of these models has been that the easier it is for a reader to identify the underlying narrative and causal structure of a text, the better his comprehension and memory for the text will be. While this assumption has received some empirical support (Thorndyke, 1977; Stein & Nezworski, 1978), the text domains to which the schemata apply have been relatively narrow. Thus the generality of particular schemata has not been determined. It is unclear, for example, whether a particular structure or schema is optimal for learning all narrative texts, or whether numerous schemata could be equally effective, depending on subtle attributes of the to-be-learned texts. Thus, as in the name-attribute-value studies discussed above, the question of whether particular schemata can be identified as optimal for learning is still unresolved.

The present study attempted to assess the overall effect of various information organizations on the learning of meaningful, naturalistic texts, and to determine if an optimal organization could be identified. If readers strongly prefer to use a single narrative schema to encode narrative passages, then a text presentation format that highlights the temporal and causal dependencies among events should produce optimal learning.

The source of materials for the experimental stimuli in Experiment 3 was newspaper stories. Several current events and feature stories were selected from the New York Times and the Los Angeles Times to serve as fact bases to be learned by subjects. The newspaper was used as a text source because news stories provide a naturalistic processing environment. People frequently read and use newspapers as an information source. Furthermore, news stories appear already organized in a standard, familiar format, so the utility of that format for learning can be contrasted with other experimentally imposed organizations. The standard structure of news stories--the presentation of "important" or "timely" information in the first paragraph and the elaboration of details and background in subsequent paragraphs--is based on conventions of journalistic style. Important information is presented first so that it will catch the reader's eye and be assimilated at a quick glance. Information is presented in generally decreasing order of importance so that if the story needs to be cut to fit a particular space on the page, the material that is cut out will be less important than what is left in. Thus, the organization of the news story is dictated by the particular requirements of that medium. However, this structure may not be optimal for the learner attempting to acquire all the facts in the passage. News stories present events out of their normal time sequence and utilize repetition and redundancy in order to establish referential connections among facts that are related but are widely separated in the story. Thus, while certain goals of the newspaper editor are met, they might decrease the overall learnability of the material.

In this chapter, the efficacy of alternative presentation formats for learning is assessed. In Experiment 3 subjects read and then recalled stories presented either in news format, standard narrative organization, or a topical organization. Measures of reading time for each passage were taken in each organizational format to provide some indication of the readability of each organization. If subjects prefer to use a narrative schema to comprehend and learn stories, then the narrative text organizations should produce faster reading times and higher recall scores than the other familiar organizations.

EXPERIMENT 3

Method

Materials. Four newspaper articles from the Los Angeles Times and the New York Times were selected for use as materials.

Each article described a set of events that occurred over a period of time and background information relevant to those events. The text of one such story is provided below as an illustration of the NEWS organization.

Iraq: News Story

> Despite having the second largest oil reserves in the Middle East, Iraq today finds itself short of cash.
>
> Civilization was cradled between its Tigris and Euphrates rivers, and the site of ancient Babylon's splendor lies 50 miles south of Bagdad.
>
> Yet the majority of Iraq's people were illiterate as late as 1973.
>
> Oil revenues have increased massively in recent years, but only a small minority have benefited, in contrast to other petroleum-producing Arab lands where spread-the-wealth has been a byword.
>
> "Our problem is management ... management from top to bottom," said Dr. Hashim Jawad, a top planning advisor to Iraq's Revolutionary Command Council.
>
> Iraq once was a sleepy, British-oriented monarchy.
>
> Red double-decker buses still churn through Bagdad's crowded streets, and what remains of the old privileged class still gathers for tea or tennis at the Alwiyah Club.
>
> To this has been added the trappings of the turbulent post-revolutionary era, such as the splendid arch which is a monument to Iraq's unknown soldier, and a still-unfinished luxury hotel.
>
> Many projects such as this have been halted in midstream and others postponed because of the lack of cash.
>
> Iraq has an estimated shortage of $600 million in oil revenues this year, out of a total expected income of around $8 billion.
>
> When the Basrah Petroleum Co., the last remaining Iraqi oil firm which still had foreign participation, was fully nationalized last spring, the former parent companies halted their purchase of oil.

> Gradually they have been coming back, with Shell and the French Petroleum Co. the first to resume.
>
> Last April, Iraq suddenly cut off all oil shipments to Mediterranean ports via a trans-Syrian pipeline.
>
> Early this year, another pipeline to the Mediterranean, which bypasses Syria and ends up in Turkey, is scheduled to become operational.
>
> This should fully make up for the capacity lost by shutting down the trans-Syrian pipeline.
>
> The present set of rulers here is the third since the 1958 revolution which ended the monarchy.
>
> That revolution set the country off on a new course, vaguely socialistic and strongly Arab nationalist.
>
> The Soviet Union is now Iraq's major friend, and Zionism is the enemy.
>
> A lack of manpower and skilled technology is the major problem.

This story concerns the declining oil revenues of Iraq, the causes for the decline, and the prospects for the future. However, many of the sentences in the story are either tangential or irrelevant to this theme. In addition, some of the sentences are repetitions or elaborations of information presented earlier.

A second version of this NEWS organization, called the CONDENSED organization, was constructed. The CONDENSED organization of each of the four newspaper stories was derived from the NEWS passages by deleting certain information from them. Deleted information was of one of three types: (1) repetition or elaboration of previously presented information; (2) background information that was irrelevant or tangential to the main point of the story; or (3) extraneous commentary on the events of the story either by the reporter or by another observer. Background information was considered to be tangential if it was neither referred to nor presupposed by subsequent statements. Thus the important information from the story was preserved, while the unimportant information was eliminated. The serial position of the remaining sentences in the story was not altered. A portion of the CONDENSED version of the Iraq story is given below.

Iraq: CONDENSED Version

> Despite having the second largest oil reserves in the Middle East, Iraq today finds itself short of cash.
>
> Iraq once was a sleepy, British-oriented monarchy.
>
> Iraq has an estimated shortage of $600 million in oil revenues this year, out of a total expected income of around $8 billion.
>
> When the Basrah Petroleum Co., the last remaining oil firm which still had foreign participation, was fully nationalized last spring, the former parent companies halted their purchase of oil.
>
> Gradually they have been coming back, with Shell and the French Petroleum Co. the first to resume.
>
> Last April, Iraq suddenly cut off all oil shipments to Mediterranean ports via a trans-Syrian pipeline....

The CONDENSED version of each news article was used to create all of the other organizations used in Experiments 3 and 4. One of these, the NARRATIVE version, was constructed by rearranging the sentences in the CONDENSED organization into a chronological sequence. Thus this passage preserved a temporal continuity in the presentation of the story's events. A portion of the NARRATIVE version of the Iraq news story is presented below.

Iraq: NARRATIVE Version

> Iraq once was a sleepy, British-oriented monarchy.

> The present set of rulers here is the third since the 1958 revolution which ended the monarchy.
>
> That revolution set the country off on a new course, vaguely socialistic and strongly Arab nationalist.
>
> The Soviet Union is now Iraq's major foreign friend, and Zionism the enemy.
>
> Iraq has the second largest oil reserves in the Middle East.
>
> Last Spring, the Basrah Petroleum Co., the last remaining Iraqi oil firm which still had foreign participation, was fully nationalized.
>
> As a result, the former parent companies halted their purchase of oil.
>
> Gradually, they have been coming back, with Shell and the French Petroleum Co. the first to resume.
>
> Then last April, Iraq suddenly cut off all oil shipments to Mediterranean ports via a trans-Syrian pipeline....

Finally, a TOPICAL version of each story was constructed by organizing the sentences from the CONDENSED version under topical subheadings. A portion of the TOPICAL passage is presented below.

Iraq: TOPICAL Version

Oil Economy

> Despite having the second largest oil reserves in the Middle East, Iraq today finds itself short of cash.
>
> Iraq has an estimated shortage of $600 million in oil revenues this year, out of a total expected income of around $8 billion....

History

> Iraq was once a sleepy, British-oriented monarchy.
>
> When the Basrah Petroleum Co., the last remaining oil firm which still had foreign participation, was fully nationalized last spring, the former parent companies halted their purchase of oil....

Politics

Last April, Iraq suddenly cut off all oil shipments to Mediterranean ports via a trans-Syrian pipeline.

This action served as an Iraqi protest of Syrian military intervention in the Lebanese civil war.

Subjects. The subjects were 60 UCLA undergraduates. They participated in the 90-minute experiment, either to satisfy a course requirement or for $5.00 pay.

Design. A Latin-square design with repeated measures was used. There were four conditions of story organization--NEWS, CONDENSED, NARRATIVE, and TOPICAL. Subjects were randomly assigned to one of four groups, with 15 subjects in each group. Each subject received one passage in each of the four organizations. The four topic stories were entitled Iraq (the story presented above), The Release of Carrillo (concerning the prison release of the Spanish Communist Party leader), Wernher Von Braun (concerning the career of the rocket scientist), and Burundi (concerning the civil war in the African country). The assignment of stories to each organization condition was counterbalanced across groups. The dependent variables were reading time and free recall of story propositions.

Procedure. Subjects were tested singly or in small groups. Subjects were given booklets containing the stories, one story per page. They were instructed to read the passages carefully because they would be asked questions about them later. They were instructed to read each passage only once and were not allowed to look back to previous stories. Each subject would read a story at a self-paced rate and then record the amount of time it took to read the story. Immediately after reading the story, subjects were instructed to write it down exactly as it appeared in wording and sentence order. However, they were told not to omit anything that they remembered simply because they could not recall its exact wording or serial position in the passage. Recalls were written on a blank sheet of paper, and unlimited recall time was provided. This read-recall procedure was repeated for each of the four stories.

Results and Discussion

In order to analyze the reading times and the recall protocols, each passage was segmented into propositions. A proposition was defined as a clause or sentence which contained an action or stative verb. Recall protocols were scored for gist reproduction of the propositions. A proposition was scored as having been correctly recalled if the relation or action in the proposition was recalled or paraphrased correctly.

The data were analyzed using a three-way analysis of variance that treated text structure, materials (story topic), and subject group as main factors. Since there was no significant interaction between structure and materials for either reading time or recall, the data were collapsed across the different story topics within each structure condition.

Reading Time. The mean reading times for each of the structure conditions are shown in the first row of Table 3.1. Mean reading time for the original news stories was the longest (157.6 seconds), while that for the other three structures was nearly equal. The effect of story organization on reading time was significant, $F(3, 168) = 22.64$, $p < .01$. Newman-Keuls tests were used to perform pairwise comparisons between the means. Reading time was significantly faster for the three other organizations than for the NEWS organization ($p < .01$ for NARRATIVE and CONDENSED; $p < .05$ for TOPICAL). No other pairwise differences were significant. That the reading time for the NEWS passages was the longest was not surprising, since these passages contained more propositions than did those in the other conditions.

To correct the reading times for the differences in passage length, each reading time was normalized by dividing it by the number of propositions in the passage. Thus for each condition, a mean reading time per proposition was obtained. These data are shown in the second row of Table 3.1. Mean reading time per proposition was actually fastest in the NEWS condition, followed by the NARRATIVE, CONDENSED, and TOPICAL conditions. This result was significant, $F(3, 168) = 3.78$, $p < .05$. The NEWS reading time was significantly faster than the TOPICAL reading time (Newman-Keuls,

Table 3.1

READING TIMES AND RECALL PERCENTAGES FOR THE TEXT ORGANIZATIONS IN EXPERIMENT 1

	Structure Type			
Factor	NEWS	CONDENSED NEWS	NARRATIVE	TOPICAL
Reading time (sec)	157.6	114.3	108.9	122.1
Reading time per proposition (sec)	2.10	2.21	2.12	2.41
Free recall (%)	18.8	25.5	24.7	24.9

$p < .05$). Again, however, the mean reading time for the CONDENSED condition did not differ from that for the structural transformations on the CONDENSED condition (NARRATIVE and TOPICAL).

Free Recall. The results for the free recall task are presented in the third row of Table 3.1, which gives the mean percentage of propositions recalled across the four passages for each text organization condition. These data were analyzed using the arcsine transformation on the percentages. Recall was lowest for the NEWS organization and higher and approximately equal for the other three organizations. The data were analyzed using the arcsine transformations of each subject's recall proportions. The effect of organization on recall was significant, $F(3, 168) = 3.13$, $p < .05$. Newman-Keuls tests showed the mean for the NEWS condition to be significantly lower than the mean for each of the other conditions ($p < .05$ for all three pairwise comparisons). The CONDENSED, NARRATIVE, and TOPICAL conditions did not differ reliably. It may be noted that while these three organizations all contained identical passage content, the NEWS passages contained additional propositions not included in the other organizations. Therefore, not only were the NEWS texts longer, but the to-be-recalled information in them was not identical to that in the other three conditions. Consequently, the analysis of the recall data was recomputed using for the NEWS condition only those propositions that occurred in the other organization conditions. When only this subset of the NEWS propositions was considered, recall was 22.0 percent. Using this scoring metric, the overall effect of organization was not significant, $F(3, 168) = 2.55$. However, a planned comparison revealed that the mean for the NEWS condition was reliably lower than the combined mean for the other conditions, $t(168) = 1.74$, $p < .05$.

The fact that recall of the NEWS passages was lower when the extraneous NEWS propositions were scored than when they were not suggests that these propositions were not well recalled. A separate analysis was performed to compare mean percentage free recall within the NEWS organization for those propositions that occurred only in the NEWS passages (i.e., the propositions that were deleted to create the CONDENSED organization) versus those propositions that occurred in all organizations. Overall, mean recall of the former propositions was 12.0 percent, while recall of the latter was 22.0 percent. This difference was significant, $t(59) = 5.42$, $p < .001$.

The finding that recall of the extraneous NEWS propositions was poorer than recall of the propositions used in the CONDENSED condition is consistent with the hypothesis that these extraneous propositions were not central to the theme of the passage. When reading the NEWS passages, subjects presumably evaluated the "importance" or "centrality" of each proposition with respect to the main theme of the story. Those that were tangential or unimportant were not processed as deeply or as carefully as the more important propositions and hence were not learned as well. This effect of propositional importance on recall has been obtained on a variety of prose materials (Kintsch, 1974; Meyer, 1975; Thorndyke, 1977).

This analysis may also be used to explain the obtained reading-time results. While recall was worst for the NEWS condition, the reading time per proposition for this condition was the fastest. It is particularly surprising that the mean propositional reading time for the NEWS condition (2.10 seconds) was faster than that for the CONDENSED condition (2.21 seconds), since these conditions were identical except for the extraneous NEWS propositions that were deleted in the CONDENSED condition. A possible explanation for this difference in reading time is that the extraneous propositions in the NEWS condition were scanned faster than those propositions that were common to the two conditions. Since subjects were reading to learn the passages, scanning time includes reading time, comprehension time, and time for elaborative processing for encoding in memory. If subjects identified the extraneous propositions as irrelevant or tangential to the theme of the passage, they could process them more superficially and hence faster than the more important ones. If the propositions common to the NEWS and CONDENSED conditions were processed at the same rate in the two conditions, while the extraneous propositions in the NEWS condition were processed faster, then the mean reading time per proposition in the NEWS condition would be faster--which, in fact, was the case.

While all structural transformations on the NEWS passages led to improved recall performance, neither the NARRATIVE nor the TOPICAL passages were read faster or recalled better than the CONDENSED passages. This result suggests that neither organization provided a preferred, familiar schema that could be used to guide comprehension and encoding of the facts embedded in those structures. Experiment 4 investigated whether this result could be replicated with another performance measure, question-answering, and with another organizational format. A TOPICAL organization was added to further highlight the narrative organization of the passages.

EXPERIMENT 4

The design of Experiment 4 was similar to that of Experiment 3 except that the NEWS and TOPICAL organizations were deleted and an OUTLINE organization was added. The OUTLINE organization was created by reorganizing the material in the CONDENSED organization into an outline format, where events were chunked into episodes according to temporal and causal associations among them. An episode thus consisted of a set of events and their consequences that were causally and topically related and that occurred together in time. The episodes were ordered chronologically. This organization is similar to that found in several psychological models of narrative story memory (Rumelhart, 1975; Thorndyke, 1977; Mandler & Johnson, 1977). In the present experiment, however, this organization was made explicit by presenting the information in the physical layout of an outline. Information was spatially arranged on the page using indentations to accentuate the hierarchical nature of the organization, explicit labels (e.g., Background, Episode 1, etc.) were given to section headings, and the text from which the outline was derived (the CONDENSED passage) was abbreviated so as to be amenable to the outline format. This alteration of the text itself required a certain amount of syntactic reduction of sentences, but no semantic alterations to the information were made. A portion of the OUTLINE organization for the Iraq text is presented below.

Iraq: OUTLINE Organization

BACKGROUND	
Iraq	Was a sleepy, British-oriented monarchy
CHRONOLOGY	
Episode 1:	1958
Event:	Revolution ended the monarchy
Result:	Country set off on a new course, vaguely socialist and strongly Arab nationalist....
Episode 2:	Last spring
Event:	The Basrah Petroleum Co., the last remaining Iraqi oil firm which still had foreign participation, was fully nationalized.
Result:	Former parent companies halted purchase of oil....
Episode 3:	Last April
Event:	Syria intervened militarily in the Lebanese civil war.
Result:	Iraq cut off oil shipments to Mediterranean ports via a trans-Syrian pipeline....

Method

Subjects. The subjects were 45 undergraduates at UCLA who participated in the experiment to satisfy a course requirement.

Design. A Latin-square design with repeated measures was used. There were three conditions of passage organization--CONDENSED, OUTLINE, and NARRATIVE--and three of the topic passages from Experiment 3 were used as materials (Iraq, Burundi, and The Release of Carrillo). Each passage had three versions, one for each of the different organization conditions. Each subject received for study and test a passage in each of the three organizations. The assignment of story topic to organization was counterbalanced across subject groups, as was the serial position of presentation of each organization.

There were two dependent variables. The first was free recall of the entire passage, as in Experiment 3, and the second was question-answering performance. For each passage, a set of 12 questions were constructed whose answers depended upon retrieving a particular detail from the passage. Answering the questions correctly required only retrieval of explicitly presented facts; no inferential processes were required. For example, one such question for the Iraq passages was "What is Iraq's major national problem?"

Procedure. Subjects were tested singly or in small groups. They were instructed to attend carefully to the passages because they would be required to recall them later. They were permitted to read each passage only once and could not look back to previous passages. After reading all three passages, subjects performed the free recall task. Recall instructions identical to those in Experiment 3 were given. Order of recall of the passages was the same as the presentation order.

The question-answering test immediately followed the free recall test. Subjects were instructed to answer the questions as well as they could, using the information they could recall from the story. Twelve questions or sentence completions were presented for each story. Questions from each story were listed on separate pages, and test order for the passages was the same as presentation order. Unlimited time for answering questions was provided.

Results and Discussion

The same procedure as that described in Experiment 4 was used to score the recall protocols. The mean percentage free recall and questions correctly answered were computed for each story in each of the three organizations. These data were analyzed separately using a three-way repeated measures analysis of variance that treated stories, structure conditions, and subject group as main effects. The results are shown in Table 3.2.

Table 3.2

FREE RECALL AND QUESTION-ANSWERING PERCENTAGES FOR THE TEXT ORGANIZATIONS AND STORIES IN EXPERIMENT 2

	Structure Type					
	CONDENSED NEWS		NARRATIVE		OUTLINE	
Story	Free Recall	Question-Answering	Free Recall	Question-Answering	Free Recall	Question-Answering
Iraq	23	36	33	52	24	42
Burundi	17	28	26	34	23	44
Carrillo	27	40	19	24	19	33
Mean	22.3	34.7	26.0	36.7	22.0	39.7

The pattern of results was similar for free recall and question-answering. Across subjects, there was a significant correlation between recall of a passage and the percent of questions correctly answered, $r = .62$, $t(133) = 9.11$, $p < .001$. For free recall, mean performance was 22.3 percent for the CONDENSED condition, 26.0 percent for the NARRATIVE condition, and 22.0 percent for the

OUTLINE condition. These differences were not significant. However, there was a significant story-by-structure interaction, $F(2, 84) = 3.70$, $p < .05$. For the question-answering task, mean percent correct was 34.7 for the CONDENSED condition, 36.7 for the NARRATIVE condition, and 39.7 for the OUTLINE condition. Again, the results by organization were not reliable, but the story-by-organization interaction was significant, $F(2, 84) = 4.35$, $p < .05$.

As in Experiment 3, none of the organizations was consistently superior for learning of the text information across stories. In fact, the interaction obtained here for both free recall and question-answering suggests that different organizations were optimal for different stories. An examination of the news stories used as materials in Experiment 4 revealed an obvious distinction among them. Both the Iraq and Burundi stories were essentially background or feature articles. They both contained a narrative chronology that provided a historical perspective for viewing the present social and political situation in the countries. Neither article contained any current, newsworthy events; rather, they focused on a broad history of events. For these stories, then, comprehension of the narrative sequence of events was critical to the theme of the passages. It is thus reasonable to presume that a presentation structure organized around the narrative chronology would facilitate learning of the text.

The Carrillo story, on the other hand, was a more typical current events story. It centered around a single event, the release of Carrillo from prison, and the multiple consequences of that event. There was much less narrative history in this story than in the others. Rather, the emphasis was on the implications in the present and in the future of the single, important action. For this story, then, a structure that emphasized the main event and its direct consequences and deemphasized the background narrative might be preferable for learning.

The ability of these hypotheses to explain the obtained interaction between structure and story was tested by computing Newman-Keuls post hoc linear contrasts on the means in Table 3.2. If a narrative structure was optimal for learning of the Iraq and Burundi stories, then free recall and question-answering performance for the NARRATIVE and OUTLINE conditions should be better than for the CONDENSED condition. For the Burundi story, the combined mean of the NARRATIVE and OUTLINE conditions was reliably greater than for the CONDENSED condition for both free recall and question-answering ($p < .05$ for both). For the Iraq story, the combined NARRATIVE and OUTLINE mean for question-answering was greater than that for the CONDENSED condition ($p < .05$). The comparison of free recall performance demonstrated the same trend but failed to achieve significance ($p < .10$).

For the Carrillo story, the hypotheses given above make the opposite prediction. That is, the CONDENSED structure, with its emphasis on the current event and situations of the story, should be superior to the NARRATIVE and OUTLINE conditions for learning. The same comparisons performed for the other stories confirmed this hypothesis. For both free recall and question-answering, performance for the CONDENSED condition was better than for the combined NARRATIVE and OUTLINE conditions ($p < .05$ for both).

GENERAL DISCUSSION

The results obtained here indicate that newspaper stories contain a considerable amount of information that is of little relevance to the reader or to the main theme of the story. A significant portion of this extraneous information is repetition of previously stated information that is necessary to establish referential identity for some new information to be presented. The repetition presumably facilitates comprehension by facilitating the integration of the new information with previously stated, or "given" information, as suggested by Haviland & Clark (1974). Since a newspaper story employs frequent shifts among the set of topics it treats, much repetition of information is required to identify the changing contexts and identify the referents for the new information to come. Other extraneous information in news stories includes incidental background information and commentary that is only tangentially related to the main point of the story. The recall data from Experiment 3 indicate that none of this extraneous information is well learned relative to the more theme-relevant information in the story. The reading-time data also suggest the possibility that this information is not processed as carefully as the more theme-relevant information. These results are consistent with the general finding in prose studies that people are more likely to learn and remember the important ideas (those central to the theme of the passage) than the unimportant ideas (Meyer, 1975; Thorndyke, 1977; Pichert & Anderson, 1977).

The recall data in the two experiments presented here complement previous work on the effects of various organizations on learning of textual material. In previous studies using texts describing the values of various attributes for a set of concepts, no clear organization emerged as optimal for learning. In the present study, texts with more substantial variation in content were used as materials. The to-be-learned information contained varying amounts of narrative history supported by background information and discussion of the consequences of the historical events. While all the passages used in this study shared this general narrative form, no single organization was found to be optimal for presenting the information in all texts.

These results argue against the existence of a single, universal schema for representing all narrative texts. The notion of a schema, as it applies to the representation of prose information, is generally interpreted to be an abstract framework or description of how narrative information can be combined to produce meaningful stories. A narrative schema, then, provides a set of constraints on how events and their consequences can be combined to produce meaningful episodes, and how a set of episodes, in turn, can be combined so they lead to a reasonable conclusion or resolution of the story. In this very general formulation, text schemata have been proposed to describe the structure of a story independent of its semantic content (Rumelhart, 1975; Thorndyke, 1977). When a person reads a story, a stored schema is presumably used to guide the comprehension of the story by imposing the constraints of the schema on the interpretation of the incoming information. When a story fits the stored schema, comprehension and retention are facilitated by the organizational and integrative benefits provided by the schema.

In the present study, however, no single organization of information was optimal for all stories. In fact, different organizations were most effective with different stories. When the story contained a narrative chronology, the organizations that emphasized the causal and temporal associations among events produced the best learning. These organizations were suggested by earlier work on schemata for narrative stories (Rumelhart, 1975; Thorndyke, 1977; Mandler & Johnson, 1977). However, when the to-be-learned story focused on a single event and its consequences, the newspaper format was superior to the narrative format as a text organization. The inability to find a consistently superior organizational form suggests one of two conclusions. Either there are no real schemata that can characterize the organization of text information, or there must be several (or many) schemata that a person can effectively use, depending on the nature of the to-be-learned information. The first conclusion seems unwarranted in light of numerous studies that indicate people have no difficulty distinguishing between stories with well-formed narrative structures and those in which normal conventions of causal and narrative organization are violated (e.g., Thorndyke, 1977; Kintsch, Mandel, & Kozminsky, 1977). While these results do not prove the existence of schemata, theories based on such structures currently provide the best theoretical account for people's ability to recognize well-formed stories. Therefore, it would seem premature to argue against schema theory on the basis of negative results from a single study.

The more reasonable conclusion from these data would seem to be that people have available a set of schemata for text organization that can be used as the content of the text warrants. For example, if the narrative information in a story is only incidental to the main point or conclusion, then a narrative schema, in which the events are assumed to lead up to and support the conclusion, may not be appropriate for encoding the story. The main point, or focus, of the story would thus be central in determining the appropriate schema for encoding the story information. In current events news stories, the main point is identified by its placement in the first sentence or paragraph of the story. On the other hand, the focus in many stories depends on the perspective of the reader. Given different perspectives on a story, different information might emerge as being central or important in development of the theme. Other researchers have begun to note circumstances in which alternative schemata may be applied to the comprehension of a story, depending on the perspective taken by the reader (Pichert & Anderson, 1977; Kozminsky, 1977). In these studies, the information retained from a passage by a reader could be influenced by biases in perspective introduced by the experimenter. These studies give credence to the notion that multiple schemata are available for use in encoding story information. For the multiple-schema theory to be useful and viable, it must be demonstrated that there are many fewer schemata than there are possible stories. That is, while there may be multiple schemata that can be used to encode story information, each one of them must be capable of representing numerous texts of its type. If the set of schemata are unable to reduce the universe of all texts to a small set of prototypical types of texts, then the theory is of little explanatory value. However, a substantial body of data have already been reported in support of a general schema for narratives, and the schema has been successfully applied to the analysis of numerous texts. It is not unreasonable to suppose that other general schemata for text organization could be identified and tested in a similar manner.

IV. INTEGRATION OF KNOWLEDGE FROM TEXT

Chapters II and III considered factors determining the acquisition of individual facts from a text. However, knowledge integration is also a fundamental component of the acquisition process. People do not simply acquire sets of unrelated facts. They integrate the facts they learn in meaningful conceptual structures. As a consequence, people can put separately acquired facts together to form new ideas. For example, a student might encounter the following two sentences at various points in a textbook chapter:

> In 1850, the Caledians rebelled because the king had declared martial law.

> The 1850 rebellion was suppressed.

By integrating the information from those two sentences, the student could respond on a subsequent examination:

> There was a rebellion in 1850 because the king declared martial law, but it was suppressed.

Knowledge integration also provides a basis for inferential reasoning. For example, a person might read in the morning newspaper:

> Mary Jones has been appointed Secretary of State.

and then hear on the evening news:

> Sam Smith has been named Special Assistant to Mary Jones.

By integrating these two news items, the person could infer:

Sam Smith has been named Special Assistant to the
Secretary of State.

Researchers have studied knowledge integration in several paradigms. Bransford and Franks (1971) provided the first experimental demonstration of knowledge integration. They showed that subjects could integrate the information in several related simple sentences to form a single, complex idea. Other researchers have demonstrated similar effects of integration of constituent ideas (Hupet & LeBoudec, 1977; James, Hillinger, & Murphy, 1977; Park & Whitten, 1977; Peterson & McIntyre, 1973). Similarly, many studies have shown that subjects can integrate several individual pairwise relations to form a single linear or partial ordering of all constituent elements (Barclay, 1973; Foos, Smith, Sabol, & Mynatt, 1976; Hayes-Roth & Hayes-Roth, 1975; Potts, 1972, 1977). A third area of research has focused on the integration of information in successive sentences based on common referents (Haviland & Clark, 1974; Clark & Haviland, 1977; Garrod & Sanford, 1976, 1977; Hupet & LeBoudec, 1977; Yekovich & Walker, in press). Finally, a number of investigators have proposed theories to account for the representation and integration of knowledge in long-term memory (Anderson, 1976; Anderson & Bower, 1973; Kintsch, 1974; Rumelhart, Lindsay, & Norman, 1972; Schank, 1976). These theories assume that the memory concepts and relations in acquired facts are represented as nodes and associations in memory. Two facts are integrated in memory if their representations share a subset of nodes and associations.

Previous studies have, in general, assumed that integration is a structural phenomenon that occurs during storage. Successively acquired facts are presumably appended to related existing knowledge representations. However, questions regarding when and how knowledge integration occurs have not been addressed.

This chapter investigates the conditions under which integration occurs. We begin with a model of knowledge integration, based on a few assumptions about memory structures and processes. Many of these assumptions appear in the previous research discussed above, and all have received previous empirical support (Hayes-Roth, 1977; Hayes-Roth & Hayes-Roth, 1977). The model provides a framework for predicting the conditions under which a given pair of facts will be integrated.

We assume that the basic units for representing facts in memory are lexical. The meanings of lexical units derive from their associative connections to other lexical units. Semantically related lexical units are presumably more closely associated than unrelated lexical units. These assumptions imply that memory representations of facts that include identical wordings can include identical subrepresentations. Memory representations of facts that include paraphrased wordings can not contain identical subrepresentations but may contain associatively connected subrepresentations. Of course, memory representations of facts that express unrelated information will have neither identical subrepresentations nor close associative connections.

Memory representations can be "activated" in two ways. They can be activated directly, by apprehension of the information they represent in an external stimulus, or associatively via excitation received from other activated memory representations. A memory representation can be activated more easily if (a) it has been activated recently; (b) it contains a subrepresentation of information that is identical to information in an external stimulus; or (c) it contains a subrepresentation that is identical to one in an activated memory representation. A memory representation is more difficult to activate if (a) it has not been activated recently; (b) it contains a subrepresentation of information that is synonymous with information in an external stimulus; or (c) it contains a subrepresentation that is associatively connected to one in an activated memory representation. Of course, a memory representation is least likely to be activated if the information it represents is unrelated to any information in an external stimulus or an activated memory representation.

We assume that when two memory representations are simultaneously active and contain identical or associatively connected subrepresentations, the two representations are integrated into a single higher-order representation. In the case of identical subrepresentations, integration effectively "superimposes" the two representations upon one another so that they share a single subrepresentation. Thus, integration eliminates representational redundancy in memory for related facts. However, the integrated representation also preserves the identities of the original constituent representations. In the case of associatively connected subrepresentations, integration establishes a direct connection between the subrepresentations reflecting the semantic relationship between them but preserving their individual identities.

Consider an example. Suppose a student studied a text describing the political organization of a particular country and encountered the following two facts:

(1) The Domestic Welfare Agency distributes information about professional options.

(2) Information about professional options is distributed by means of computer terminals.

Facts (1) and (2) share a common topic (the distribution of information about professional options) and they present complementary details regarding that topic (that the information is distributed by the Domestic Welfare Agency and that it is distributed by means of computer terminals). The student could integrate the two facts as a single "idea" that included all of the information:

(3) The Domestic Welfare Agency distributes information about professional options by means of computer terminals.

According to the assumptions outlined above, simply learning (1) and (2) does not guarantee that they will be integrated. Successful integration requires the simultaneous activation of their representations. This activation could occur directly because the two facts contain identical wordings of common information, i.e., "distributed information about professional options." Let us consider the memory representations that would result from either the success or failure of the integration process.

If the (1) and (2) representations were not simultaneously activated, they would have independent representations, as shown in Fig. 4.1. For purposes of illustration, we have made a few arbitrary assumptions regarding the structural details of the individual fact representations. Figure 4.1 illustrates two important points. First, the representations preserve the lexical constituents of the input facts. Second, the two fact representations remain unintegrated; that is, they share no common subrepresentations, and no direct associations connect them. (In a complete memory, indirect associations would connect these fact representations via other representations defining their lexical constituents. We have omitted these from Fig. 4.1 for simplicity.)

If the (1) and (2) representations were simultaneously activated, they would be integrated in a single higher-order representation, as illustrated in Fig. 4.2. The two fact representations have upon one another, so that they share a common subrepresentation of the shared information, "distributes information about professional options." Note, however, that the higher-order representation does not simply incorporate the constituent representations, obscuring the distinction between (1) and (2). Instead, the higher-order representation preserves the distinction, as indicated by the solid and broken lines in Fig. 4.2.

Now consider the case in which the student encountered the following instead of fact (1):

(1') The Domestic Welfare Agency provides career counseling.

Integration of (2) with the previously learned (1') would depend upon simultaneous activation of the two representations. Because the two

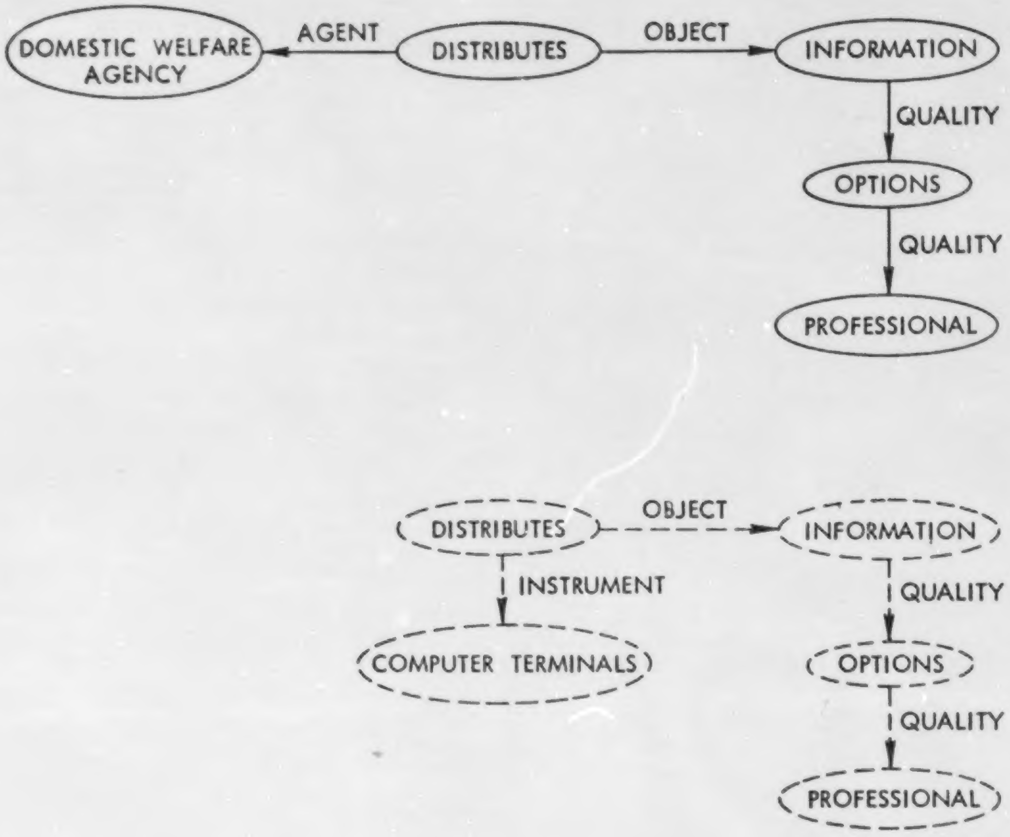
DOMESTIC WELFARE AGENCY
AGENT
DISTRIBUTES
OBJECT
INFORMATION
QUALITY
OPTIONS
QUALITY
PROFESSIONAL
DISTRIBUTES
OBJECT
INFORMATION
INSTRUMENT
QUALITY
COMPUTER TERMINALS
OPTIONS
QUALITY
PROFESSIONAL

Fig. 4.1--Non-integrated assemblies representing complementary facts (1) and (2)

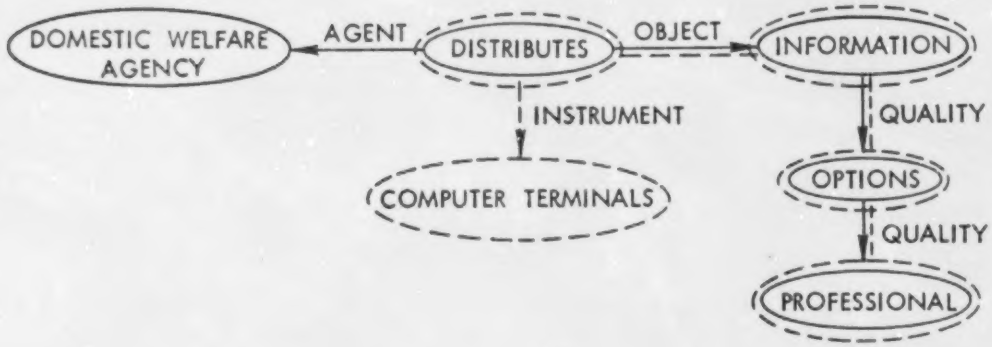
DOMESTIC WELFARE AGENCY
AGENT
DISTRIBUTES
OBJECT
INFORMATION
INSTRUMENT
QUALITY
COMPUTER TERMINALS
OPTIONS
QUALITY
PROFESSIONAL

Fig. 4.2--Integrated assembly representing complementary facts (1) and (2) facts contain paraphrases of common information ("distributes information about professional options" versus "provides career counseling") this activation could occur only associatively. In other words, activation of (1') during input of (2) would require associative "chaining" based on the semantic relationship between the two paraphrases.

If the (1') and (2) representations were activated simultaneously, they would be integrated in a single higher-order representation, as illustrated in Fig. 4.3. Because the common information is paraphrased in the two facts, the two representations have no sharable subrepresentations and cannot be superimposed upon one another. Instead, integration is accomplished by encoding the semantic equivalence of the common information as an equivalence relation between the two synonymous subrepresentations.

As discussed in these examples, integration of (1) (or 1') and (2) presumably occurs only if the two fact representations are

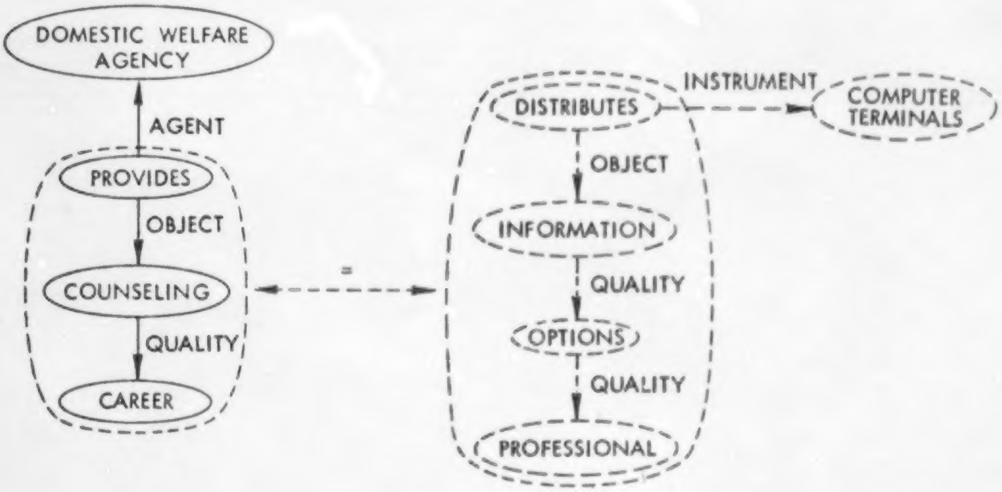
DOMESTIC WELFARE AGENCY
AGENT
PROVIDES
OBJECT
COUNSELING
QUALITY
CAREER
=
INSTRUMENT
DISTRIBUTES
COMPUTER TERMINALS
OBJECT
INFORMATION
QUALITY
OPTIONS
QUALITY
PROFESSIONAL

Fig. 4.3--Integrated assembly representing differently worded complementary facts (1') and (2) activated simultaneously. Two factors influence whether or not simultaneous activation occurs. The first factor is the recency of activation of the (1) representation. As the time since the most recent activation of (1) increases, the probability of its activation when (2) is encountered decreases. Therefore, the probability that integration will occur decreases. The second factor is the correspondence between information in the two facts. Representations of facts containing identical wordings of common information, such as (1) and (2), can activate one another directly. Representations of facts containing paraphrases of common information, such as (1') and (2), must activate one another associatively. Therefore, simultaneous activation and successful integration are more likely in the former case than in the latter.

The following experiments investigated these predictions. In these experiments, the two factors of interest were operationalized as binary variables. Thus, related facts occurred either consecutively in a single story or in two different stories. Similarly, common information in related facts was either worded identically or paraphrased. Of course, if one could quantify the proximity between related facts or the similarity of their wordings, one could presumably predict the magnitudes of the effects. However, no effort was made to quantify these variables, and therefore only the qualitative nature of their effects will be evaluated.

EXPERIMENT 5

Experiment 5 tested these predictions in a task requiring subjects to integrate two facts in memory. The facts provided information necessary to fill different slots in a single case frame, as illustrated by (1) or (1') and (2) above. Subjects were presented pairs of related facts embedded in meaningful stories and were then tested on their ability to integrate the related facts.

A matching test measured subjects' ability to identify pairs of case fillers from different facts that shared the same case frame. Matching of case fillers was performed in the absence of explicit case-frame cues (information common to the related facts). For example, subjects were tested on their ability to identify "Domestic Welfare Agency" and "computer terminals" as the agent and instrument from a single case frame. The case frame itself ("distributes information about professional options") was not given on the test.

In performing the matching task, subjects presumably could use a given case filler to activate the memory representation in which it occurred. If that representation were part of an integrated representation of all facts involving the case frame, its activation would provide access to all other associated case fillers, that is, those that were appropriate matches. If the activated representation were not part of an integrated representation, the subject would have to use the activated case-frame subrepresentation as a cue to associatively activate a second knowledge representation in which the same or a semantically equivalent case frame occurred in order to locate other associated case fillers. Because of the additional processing required in the latter case, retrieval of case fillers should be less likely and performance on the matching test should be worse.

On another task, subjects were cued with case frames they had seen previously (i.e., information common to the related facts) and were then asked to identify the pairs of case fillers occurring in that case frame from a long list of alternatives. In performing this multiple-choice task, subjects presumably used the case-frame cue to activate representations in which it occurred. If the case-frame cue occurred in an integrated representation, activating it would also activate all of the associated case fillers. Subjects could then use this information to select the appropriate responses to the case frame cue from the list of alternatives. If the case frame cue occurred in two independent representations, both of them would have to be activated in order to retrieve all associated case fillers. Again, because of the extra processing required in the latter case, activation of all case fillers should be less likely and performance on the multiple choice test should be worse than when the fillers were stored in a single integrated representation.

Two independent variables were manipulated to influence the probability that related facts would be integrated. Related fact pairs such as (1) and (2) occurred either consecutively within a single story or in two different stories. According to the assumptions outlined above, consecutive occurrence of the two facts in a single story should facilitate integration. Thus, performance on the matching and cued recall tasks should be better in the one-story condition than in the two-story condition. Pairs of related facts also varied in the wordings of common information. Common information was either worded identically or paraphrased, as illustrated by the pairing of either (1) or (1') with (2) above. According to the assumptions outlined above, identical wording of the common information should facilitate integration. Thus, performance on the matching and cued-recall tasks should be better in the identical-wording condition than in the paraphrase condition.

The assumptions also predict an interaction between the effects of number of stories and wording of the common information. In the one-story condition, consecutive presentation of related facts should facilitate both direct and associative activation. This facilitation may be strong enough that identical wordings of common information provide no additional advantage. Therefore, the wording manipulation should have a relatively small effect in the one-story condition. In the two-story condition, on the other hand, both direct and associative matches are less likely to occur. In this condition, identical wordings of common information should facilitate direct activation and, hence, integration. Thus, while the effect of the wording manipulation should be relatively small in the one-story condition, it should be relatively large in the two-story condition.

Method

Materials. Three sets of meaningful stories were constructed. Each set consisted of three stories about a different mythical country, and each story within a set was about a different aspect of that country. Six pairs of related facts and six unpaired and unrelated filler facts were equally distributed among the three stories in each set. Each pair of related facts contained information necessary to fill different slots in a single case frame, as illustrated in (1) and (2) above. Four versions of these materials were used in Experiment 5. In the one-story condition, the two facts constituting a related pair occurred consecutively in a story. There were two such pairs in each story within the set of passages for a given country. In the two-story condition, the constituent facts in a pair occurred in different stories within a set. Two of the related pairs had constituents in Stories 1 and 2 in the set, two had constituents in Stories 1 and 3, and two had constituents in Stories 2 and 3.

In both the one- and two-story conditions, each pair of related facts had either the identical wording or a paraphrase of common information, as shown above by the alternative pairings of (1) or (1') with (2). As an illustration of these materials, one set of stories exemplifying the identical-wordings two-story condition is presented below.

Brownland 1

In Brownland, the work of the government is divided among several different bureaucratic agencies. Some of the agencies and their responsibilities are given below. The National Intelligence Group collects data regarding the international superpowers. The Navy attacks enemies of Brownland. The Board of Banking studies supply and demand fluctuations in order to prevent fiscal crises. The Royal Knowledge Society monitors scientific investigations in universities. The Internal Guard uses negotiations to deal with civil riots. The Domestic Welfare Agency distributes information about professional options to all citizens.

Brownland 2

Government activities in Brownland are undertaken with particular purposes in mind. A representative sample of

> activities and purposes is given below. The movement of citizens within Brownland is reported to the Statistics Department in order to minimize census taking difficulties. Spying operations are undertaken primarily to evaluate the likelihood that Brownland will be invaded. The government collects data regarding the international superpowers in an effort to anticipate major disruptions. Scientific investigations in universities are monitored so that important findings can be made available to the government. The state keeps track of the wealth of individual citizens in order to facilitate economic planning. County agents maintain permanent files of all violations of the law so that repeat offenders can be punished.

> Brownland 3
>
> The Brownland government makes use of various kinds of equipment and personnel in carrying out its functions. Some of these are described below. Social workers are used to insure that children are given adequate home environments in order to promote an egalitarian society. The vice squad uses electronic surveillance equipment to detect crime in the streets at night. Long range missiles are used to attack enemies. Spying operations utilize paratroopers. Information about professional options is distributed by means of computer terminals. The state keeps track of the wealth of individual citizens by means of ID cards.

Subjects. Sixty-four UCLA undergraduates participated in the one-hour experiment. Subjects were either paid $2.50 or given course credit for participation.

Design. A 2 x 2 between-subjects factorial design was used. The location of pair constituents (one-story versus two-story conditions) was crossed with wording of the constituents (identical versus paraphrase) to produce four experimental conditions. Subjects were assigned randomly to one of the four groups.

Procedure. Subjects were tested in groups. Each subject was given a booklet containing the experimental stories and tests. Subjects' progress through the booklets was self-paced.

Subjects studied and were tested on each set of three stories as follows. They read the first story carefully; then they performed a cued recall test for facts from the story; then they looked back at the story to check their answers and study any facts they missed. This read-recall-check procedure was repeated for each of the three stories in the set.

Then subjects were given a matching test. Two lists were presented, each of which contained the case fillers from all of the facts in the stories. For example, the lists for Brownland included the following two items: "Domestic Welfare Agency" and "computer terminals." One of the lists had a blank space preceding each item; the other had the items numbered. The subject's task was to indicate which items occurred in the same case frame by writing the numbers of items from the numbered list in the blanks preceding the corresponding constituent fillers in the other list.

After completion of the matching task, subjects were given a multiple-choice task. Subjects were cued with the subsets of information common to facts in a related pair (e.g. "distributes information about professional options"). Only one of the two wordings was used as a cue. The subject's task was to select from the list of all case fillers that had occurred those that were appropriate for each of the cues (e.g. "Domestic Welfare Agency" and "computer terminals").

This entire procedure was repeated for each set of stories.

Results

The probabilities of correct responses on the matching and multiple-choice tests are shown in Table 4.1. Since errors [illegible] the tests could be produced either by memory failure for individual facts or by failure to integrate related facts, test performance was considerably lower than 100 percent.

The results of the matching test are shown in the upper part of Table 4.1. The entries give the probabilities of a correct match between case fillers associated with the same case frame for the four experimental conditions. Performance was better when in the one-story condition than in the two-story condition, $\underline{F}(1,60) = 8.32$, $\underline{p} < .01$.

Table 4.1

PERFORMANCE ON THE MATCHING AND MULTIPLE-CHOICE TESTS

	Wording of Common Information in Related Facts			
	Identical	Paraphrase	Identical	Paraphrase
Location of Related Facts	Correct Matches (%)		Correct Choices of Pairs of Items (%)	
One story	.45	.40	.57	.52
Two stories	.35	.19	.44	.25

In addition, performance was better in the identical-wording condition than in the paraphrase condition, F (1,60) = 4.05, $p < .05$. However, the latter difference was significant only in the two-story condition (for the one-story condition, $t(30) = .62$; for the two-story condition, $t(30) = 2.53$, $p < .01$).

The results of the multiple-choice test are shown in the lower part of Table 4.1. These entries give the probabilities of correct identification of both case fillers associated with the case-frame cue for the four experimental conditions. These data are comparable to the results of the matching test. Performance was better in the one-story condition than in the two-story condition, $F(1,60) = 14.20$, $p < .01$; and it was better in the identical-wording condition than in the paraphrase condition, $F(1,60) = 5.05$, $p < .05$. Again, this difference was significant only in the one-story condition (for the one-story condition, $t(30) = .62$; for the two-story condition, $t(30) = 2.78$, $p < .01$).

Multiple-choice data from the paraphrase condition (see Table 4.2) were analyzed further. In the multiple-choice task, half of the case fillers in the paraphrase condition had been presented originally with the case-frame cue given in the test, and half had been presented with a paraphrase of this cue. Table 4.2 shows that, overall, individual case fillers were more likely to be identified if they had been presented originally with the test cue (.69) than if they had been presented originally with a paraphrase of the test cue (.52), $F(1,30) = 24.66$, $p < .001$. This effect was greater in the two-story condition than in the one-story condition, $F(1,30) = 4.01$, $p < .05$.

Table 4.2

PERCENT CORRECT CHOICES OF INDIVIDUAL ITEMS

Location of Related Facts	Wording of Cue	
	Identical	Paraphrase
One story	.72	.66
Two stories	.59	.45

Discussion

These results are consonant with the predictions discussed above. When related facts, such as (1) and (2), occur consecutively in a story, integration of the two representations is highly probable. As predicted, the wording manipulation had only a small effect when related facts occurred together. Subjects performed well on both tests, regardless of whether common information in related facts had identical or paraphrase wordings. This is consistent with our assumption that associative activation, as well as direct activation of the (1) and (2) representations, is facilitated by recent prior activation of (1). Additional evidence that this integration occurred was obtained from the analysis of matching test performance within the paraphrase condition. If the related facts were integrated, subjects should have been able to identify case fillers originally presented with the test cues and fillers originally presented with paraphrases of the cues. In fact, performance was quite good on both kinds of case fillers, and the difference between them was small.

When related facts occur in two different stories, the (1) representation is weaker during input of (2) than it is when the facts occur together. Thus, integration of the two facts is less likely than in the one-story condition. This was reflected in the observation that subjects performed substantially worse in the two-story condition than in the one-story condition on both the matching and multiple-choice tests. As predicted, the wording manipulation produced a large effect in this condition. Subjects performed better on both tests when common information had identical wordings in the two facts than when it was paraphrased. This is consistent with our assumption that associative activation is less likely to succeed than direct activation. Additional evidence on this point derives from the analysis of matching test performance within the paraphrase condition. If related facts are not integrated, subjects should be more likely to identify case fillers presented originally with the test cues than fillers presented originally with paraphrases of the cues. Performance was worse on both kinds of items than in the same-story condition (where many pairs of related facts in the paraphrase condition were presumably integrated). More importantly, case fillers presented originally with the test cues were more likely to be identified correctly than fillers presented originally with paraphrases of the cues.

The observed interaction between story and wording conditions is particularly noteworthy. Many researchers have supposed that both lexical and more abstract semantic codes exist in memory, but that the former fade rapidly, while the latter persist (Dooling, Christiaansen, & Keenan, 1975; Fillenbaum, 1966; Sachs, 1974). This view predicts that the effect of the wording manipulation should decrease as the temporal interval between presentation of related facts increases. However, exactly the opposite result was obtained here.

EXPERIMENT 6

Some researchers have assumed that the integration of related facts into a unified memory representation obscures memory of the unique occurrences of the facts. For example, Bransford and Franks (1971) observed that subjects could not discriminate OLD sentences from NEW sentences that were consistent with the information in the OLD sentence. They concluded: "Individual sentences lost their unique status in memory in favor of a more holistic representation of semantic memory" (p. 348). However, subsequent research (Katz, Ateson, & Lee, 1974; Katz & Gruenewald, 1974) suggested alternative interpretations of these data. In particular, the work of Bransford, Barclay, & Franks (1972) and James, Hillinger, & Murphy (1977) indicated that people retain knowledge of originally presented facts even when those facts are integrated with others in memory.

As discussed above, we agree with the latter assumption that integration of related facts does not completely obscure their separate identities. When related facts having paraphrased wordings are integrated, the preservation of wording information in memory distinguishes them, as shown in Fig. 4.3. Even when similarly worded facts are integrated by the sharing of common representations, however, separate traces distinguish the facts, as shown in Fig. 4.2. The model can accommodate a small percentage of false recognitions of "integrated" facts (i.e., combinations of related facts) by assuming that the distinctive traces or tags encoding separately recorded facts might occasionally deteriorate, while the rest of the information persisted. Thus, the theory predicts that there should be fewer false recognitions of "integrated" facts than correct recognitions of facts that were actually presented. This discriminability should hold regardless of whether the common information in the related facts has identical or paraphrase wordings.

Experiment 6 tested this hypothesis for the materials and conditions of Experiment 5. Subjects were given a combined recognition-verification test containing OLD items, which had actually been presented; INFERENCES, which integrated the information in two separately presented but related facts; and NEW items, which contained concepts and relations from presented facts but combined them inappropriately. For each test sentence, subjects judged whether the sentence was OLD (had been presented originally) or NEW (had not been presented originally). For items judged to be NEW, subjects indicated whether the sentence was TRUE (stated information that was true in the passage although not expressed in a single sentence) or FALSE.

We assume that either an OLD or a NEW-TRUE response to a test item implies that the subject has learned the information in the item. Thus, an OLD or NEW-TRUE response to an OLD item implies that the subject has learned the item. An OLD or TRUE response to an INFERENCE implies that the subject has integrated the facts necessary to derive the INFERENCE.

We assume that an OLD response to a test item implies that in addition to having learned the information in the item, the subject perceived an effectively perfect match between the test item and its memory representation. Thus, an OLD response to an OLD item implies that its memory representation has remained relatively intact. An OLD response to an INFERENCE implies that any trace encoding the individual identities of the constituent facts has deteriorated, while the memory representation of the remainder of the information has persisted.

The prediction can be restated in terms of OLD versus TRUE response probabilities for OLD items and INFERENCES. If constituent facts retain representational integrity even when integrated, subjects should be able to discriminate presented facts from INFERENCES in all conditions. That is, relatively few OLD responses should occur for INFERENCES, and the probability of an OLD response should be substantially lower for INFERENCES than for OLD test items. Integrated knowledge of related facts should, instead, be exhibited as high rates of NEW-TRUE reponses to INFERENCES. If, on the other hand, constituent facts lose their identities in integrated memory representations, subjects should produce many OLD responses to INFERENCES on a recognition test. When integration is very likely to occur (in the one-story, same-wording condition, in particular), such false-alarm rates for INFERENCES should approach hit rates for OLD items. Relatively few NEW-TRUE responses should occur for either item type.

Method

Materials. The materials were the same sets of stories used in Experiment 5, with the modification that new filler sentences were constructed that specified two filled cases in the sentence case frame, rather than one filled case, as in the constituent facts of related pairs. These filler sentences occupied their same serial positions in the passages as in Experiment 5, so that each passage contained some sentences with one case specified and some sentences with two cases specified.

Three types of items were constructed for the recognition-verification test. The OLD items for each set of three stories comprised the six filler sentences from the three stories. Six INFERENCES were constructed by combining each of the six pairs of separately presented but related facts into single sentences. For example, one such item constructed from the materials used in Experiment 5 was "The Domestic Welfare Agency distributes information about professional options using computer terminals." These sentences could be correctly classified as NEW and TRUE. Six false NEW sentences were constructed by inappropriately integrating the information from two separately presented sentences. Thus all test items specified filler information for two slots in a case frame. There were 54 test items in all, 18 for each set of stories.

Subjects. Sixty-four UCLA undergraduates participated in the experiment, either for payment of $2.50 or to fulfill a course requirement.

Design and Procedure. The four experimental conditions were identical to those in Experiment 5 (identical versus paraphrase wordings of related facts crossed with one or two stories). Subjects were randomly assigned to one of the four conditions. In each condition, there were three types of test items: OLDs, NEWs, and INFERENCEs.

Each subject worked individually with a booklet containing the stories. Subjects studied each set of three stories, using the study-recall-check procedure described for Experiment 5. Then they were given a combined recognition-verification test. The 18 test items were presented in random order. Subjects indicated whether each test item was OLD (had occurred exactly as stated in the studied stories) or NEW. If an item was judged to be NEW, the subject also indicated whether it was TRUE or FALSE. A TRUE response meant the subject believed the fact gave true information from the stories, even though the sentence had not been presented explicitly.

Results

For each subject in each condition, the probability of responding OLD and the probability of responding OLD or NEW-TRUE were tabulated for OLD and INFERENCE test items. (The latter probability was simply the sum of the probability of responding OLD and that of responding NEW-TRUE.) Each of these probabilities was corrected for guessing, using a variation of the high-threshold correction. The following corrections were used:

$$\text{Corrected } P(\text{OLD} \mid \text{OLD}) = \{P(\text{OLD} \mid \text{OLD}) - P(\text{OLD} \mid \text{NEW})\}/\{1 - P(\text{OLD} \mid \text{NEW})\}$$

$$\text{Corrected } P(\text{OLD} \mid \text{INFERENCE}) = \{P(\text{OLD} \mid \text{INFERENCE}) - P(\text{OLD} \mid \text{NEW})\}/\{1 - P(\text{OLD} \mid \text{NEW})\}$$

$$\text{Corrected } P(\text{OLD or TRUE} \mid \text{OLD}) = \{P(\text{OLD or TRUE} \mid \text{OLD}) - P(\text{OLD or TRUE} \mid \text{NEW})\}/\{1 - P(\text{OLD or TRUE} \mid \text{NEW})\}$$

$$\text{Corrected } P(\text{OLD or TRUE} \mid \text{INFERENCE}) = \{P(\text{OLD or TRUE} \mid \text{INFERENCE}) - P(\text{OLD or TRUE} \mid \text{NEW})\}/\{1 - P(\text{OLD or TRUE} \mid \text{NEW})\}$$

The data for OLD responses and NEW-TRUE responses were analyzed separately. In each case, the data were submitted to an analysis of variance that treated wordings, number of stories, and items (OLD or INFERENCE) as main effects.

The results are given in Table 4.3. The top half of the table presents the proportions of OLD responses given to OLD and INFERENCE test items in each of the four conditions (identical versus paraphrase wording crossed with one versus two stories). In all four conditions, subjects made fewer OLD responses to INFERENCES than to OLD items. The main effect of items was significant, $F(1,36) = 85.61$, $p < .001$. The probability of an OLD response was greater in the one-story (.57) than in the two-story condition (.42). This main effect of number of stories was also significant, $F(1,36) = 19.21$, $p < .001$.

The bottom half of Table 4.3 presents the proportions of OLD or NEW-TRUE responses given to OLD items and INFERENCES in each condition. Subjects made more OLD or NEW-TRUE responses to both OLD items and INFERENCES in the one-story condition than in the two-story condition. There was a main effect of number of stories, $F(1,36) = 10.97$, $p < .01$, but no main effect of item type (OLD versus INFERENCE $F(1, 36) = 0.584$. However, there was a significant item-type - by - wording (identical versus paraphrase) interaction, $F(1,36) = 5.66$, $p < .05$. That is, while there was no difference between OLD or NEW-TRUE response probabilities to OLD versus INFERENCE items in the identical wording condition, these probabilities were higher for OLD items than for INFERENCES in the paraphrase wording condition.

Table 4.3

RESPONSE PROBABILITIES ON RECOGNITION TEST

	Same Story		Different Stories	
	Wording of Common Information in Related Facts			
Item Type	Same	Paraphrase	Same	Paraphrase
P(OLD)				
OLD	.77	.84	.58	.60
INFERENCE	.28	.38	.28	.22
P(OLD or TRUE)				
OLD	.87	.86	.64	.72
INFERENCE	.89	.79	.73	.58

Discussion

The absence of any main effect of item type on probability of an OLD or NEW-TRUE response indicates that, in general, subjects successfully integrated related facts as often as they learned OLD items. Thus, if integration obscures the identities of constituent facts, subjects should have responded OLD to INFERENCES as often as to OLD test items. However, as predicted, substantially higher probabilities of OLD responses were observed for OLD items than for INFERENCES. Subjects apparently retain information about the individual identities of the facts they study even when those facts are integrated in memory.

It might be argued that in some conditions (for example, the paraphrase wording condition), subjects did not integrate related facts as often as they learned OLD items. This would also produce the observed effect. However, in the one-story, identical-wording condition there was clearly no difference in the probability of integrating related facts and the probability of learning OLD items (i.e., no difference between probabilities of OLD or NEW-TRUE responses to OLD items and INFERENCES). Yet it was in this condition that the largest difference between probabilities of OLD responses to OLD items versus INFERENCES was observed.

The results of Experiment 6 provide additional support for some of the predictions supported in Experiment 5. The OLD or NEW-TRUE responses to INFERENCES indicated that subjects were more likely to integrate related facts that occurred in one story than those that occurred in two different stories. They also indicated that subjects were more likely to integrate related facts if the common information in those facts had identical wording than if it was paraphrased.

EXPERIMENT 7

Experiments 5 and 6 investigated integration of facts that were constituents of higher-order knowledge units. The inferential process enabled by integration was essentially concatenation of two knowledge structures.

People also integrate facts having other kinds of relationships to enable more sophisticated inferential processes. Consider, for example, the following two facts:

(4) Albert Profiro hated all dictators.

(5) King Egbert was a dictator.

Integration of these two facts would provide a valid basis for inferring:

(6) Albert Profiro hated King Egbert.

This experiment investigated integration of facts such as (4) and (5) above and the influence of integration on subjects' performance of deductive reasoning, using the facts as premises. Subjects studied several pairs of stories containing pairs of related facts. Each pair of related facts could be used to support an inference. After studying each pair of stories, subjects verified a set of true and false inferences.

Consider what the subject might be doing in order to verify an inference such as (6) above. Correct verification of a true inference requires simultaneous consideration of two studied facts (for example, facts (4) and (5)). Presumably, subjects would attempt to use the information in a test item to cue retrieval of facts from memory to support it. For example, the subject could use the information "Albert Profiro hated" to cue retrieval of (4) and use the information "King Egbert" to cue retrieval of (5). Given successful retrieval of (4) and (5), the subject must reason across the two premises to validate (6). (In the case of a false test item, the subject will presumably fail to retrieve any pair of facts that can function as premises in the validation of the test item and will thus respond FALSE.)

Figures 4.4 and 4.5 illustrate non-integrated and integrated representations of (4) and (5). Again, the details of individual fact

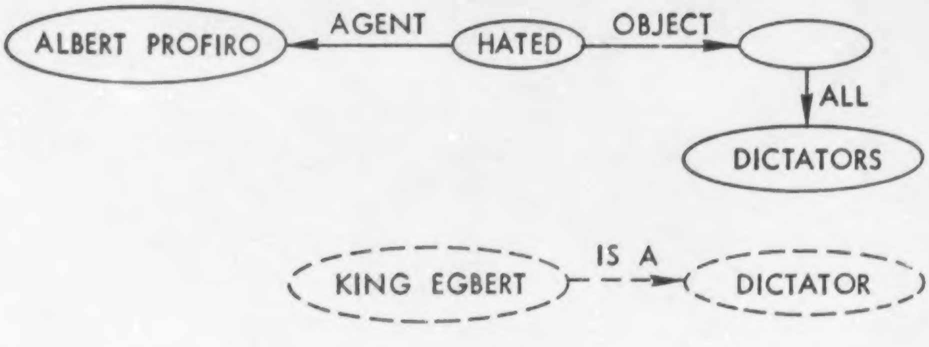
ALBERT PROFIRO
AGENT
HATED
OBJECT
ALL
DICTATORS
KING EGBERT
IS A
DICTATOR

Fig. 4.4--Non-integrated assemblies representing related facts (4) and (5)

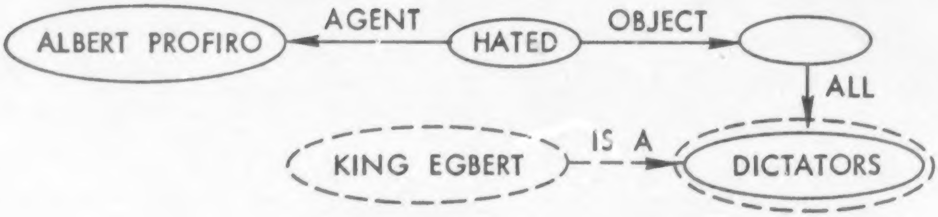
ALBERT PROFIRO
AGENT
HATED
OBJECT
ALL
KING EGBERT
IS A
DICTATORS

Fig. 4.5--Integrated assembly representing related facts (4) and (5) representations are arbitrary and should be disregarded. The important aspects of Figs. 4.4 and 4.5 are the structural relationships between the two fact representations. The non-integrated representations in Fig. 4.4 have no shared structural components. The integrated representations in Fig. 4.5 share a common subrepresentation of "dictator."

If the facts necessary to validate a test item have non-integrated memory representations, the subject must retrieve each of the facts independently. That is, the subject must use the information "Albert hates someone" to retrieve (4) and then, independently, use the information "King Egbert" to retrieve (5). If, on the other hand, the two facts have integrated memory representations, activation of either one of them entails activating the other. That is, the subject can use the information "Albert hates someone" to retrieve (4) and (5). Alternatively, the subject can use the information "King Egbert" to retrieve (4) and (5). Thus, integrated memory representations of related facts such as (4) and (5) can facilitate inferencing based on those facts in two ways: First, less processing is required to activate one integrated representation than to activate two independent representations. Second, two independent cues are available to activate the integrated representation, compared to the single pair of cues available to activate the two independent representations.

Two independent variables were manipulated in this experiment. Common information in the pairs of stories subjects studied, and particularly in the pairs of facts necessary to verify test inferences, had either identical or paraphrase wordings. Pairs of facts in the identical-wording condition are illustrated by (4) and (5) above. Pairs of facts in the paraphrase condition are illustrated by (4) and (5'):

(5') King Egbert was an autocrat.

According to the assumptions outlined above, integration of the two facts can occur in either condition. (Integration of (4) and (5') would require encoding of the equivalence relation between "dictator" and "autocrat," as illustrated in Fig. 4.3 for the equivalence relation between "provides career counseling" and "distributes information about professional options.") However, identical wordings of the common information should facilitate integration. Therefore, subjects should perform better on the inference test in the identical-wording condition than in the paraphrase condition.

The second independent variable was retention interval. Subjects performed the inference test after either 0 or 30 minutes. If the wording manipulation produces the assumed effect on memory representations, relative performance in the identical-wording versus paraphrase conditions should be comparable for both retention intervals.

Method

Materials. Four pairs of meaningful stories were constructed. All of the stories were about the mythical country Morinthia. Each individual story was about a different topic, but the stories within a pair were about related topics. The four pairs of topics were (1) The First Morinthian Revolution, and The Imprisonment of Albert Profiro; (2) Religious Customs and Beliefs in Morinthia, and (Religious Overtones During) the Fever Epidemic; (3) The Marriage of Princess Isadora, Successor to the Throne of Morinthia, and The Romance Between Princess Mathilde and Basil; and (4) The Provincial Lifestyle in Morinthia, and The Home of the Caledian Ambassador.

Each pair of stories included four pairs of related facts, such as (4) and (5) or (5') above. Each pair of related facts contained the information necessary to support a particular inference not explicitly stated in either story (such as (6) above). Related stories and, in particular, the related facts within the stories had either the identical or paraphrase wordings of common information, as illustrated above by the alternative pairings of (4) with (5) or (5'). As an illustration of these materials, one set of stories exemplifying the identical-wordings condition is presented below.

The First Morinthian Revolution

The Spring Episode was the first revolution in Morinthia. The outbreak occurred shortly before dawn on April 17, 1843. The revolution was undoubtedly caused by the tyranny imposed upon the Morinthian people by King Egbert, the dictator. For months, Egbert had extracted half of all the earnings of the people. However, the immediate cause of the outbreak appeared to be a minor crime committed several days earlier. A peasant had poached several chickens from the royal henhouse to serve at his daughter's wedding. It seemed a minor offense to the people, but in Morinthia, everyone who disobeyed the law was punished severely. The peasant was branded one of the king's enemies and thrown into prison. The Morinthian prison was populated exclusively by the king's enemies. The townspeople were thrown into a frenzy at the severity of the sentence. Even those who swore loyalty to Egbert joined the crowds demanding freedom for the peasant. The crowds stormed the palace. An effigy of the king was burned. Egbert commanded them to respect his authority and disperse at once. In the end, Egbert called out his guards and martial law was imposed. So ended the first Morinthian revolution, all of which were doomed to failure.

The Imprisonment of Albert Profiro

The Curfew Episode was the second revolution in Morinthia. It provided the setting for several important events in the life of Albert Profiro, a young Morinthian tradesman. The outbreak occurred on March 22, 1844, the day after a group of youths were discovered to have disobeyed the curfew law. The law had been a source of friction between the townspeople and the government for some time. The people welcomed the opportunity to flood the streets, throwing stones and damaging property. Albert took it upon himself to try to calm the people. Although Albert hated all dictators and their governments, he hated anarchy in the streets even more. So he positioned himself on a platform in the center of the town square and called upon the people to return to their homes. Unfortunately, when the royal soldiers arrived, they only saw a young man shouting to the crowd and assumed he was responsible for the riot. Albert was arrested and thrown into prison. Although Albert spent three bitter years in prison, his experience brought some good with it as well. It was in prison that Albert met Anastacia DeVille, whom he subsequently married.

Subjects. Twenty-six UCLA undergraduates participated in the two-hour experiment. Subjects were either paid $2.50 or given course credit for their participation.

Design. A 2 x 2 factorial design was used. The wording of the common information in the two stories (identical or paraphrase) was crossed with retention interval (0 or 30 minutes) to produce four conditions. The wording manipulation was a within-subject factor. The retention-interval manipulation was a between-subjects factor, with 11 subjects in the 0-minute retention-interval condition and 15 subjects in the 30-minute retention-interval condition.

Procedure. Subjects were tested in groups. Each subject was given a booklet containing the experimental stories and tests. Subjects' progress through the booklets was self-paced. Intentional learning instructions were given, including the warning that an inference test would be given.

Subjects studied and were tested on each of the three stories as follows. They read a pair of related stories carefully, attempting to learn as much as possible. Then they were given a verification test. On this test, TRUE items were defined as those that could be proved true, given the information in the stories. FALSE items were defined as those that could not be proved true, given the information in the stories. There were four TRUE and four FALSE items, ordered randomly, on the test following each pair of stories. This study-test procedure was repeated for each of the four pairs of topically related stories.

After subjects had studied and been tested on all stories, they were given a final test of the inferences in syllogism form. On this test, each inference was immediately preceded by the two facts that presumably determined its validity. Subjects simply indicated whether or not each inference followed logically from the two facts that preceded it.

Results and Discussion

The analysis of performance on the verification task included only those items to which the subject had responded "correctly" on the final syllogism test. That is, inferences for which a subject could not perform the necessary reasoning correctly, given the premises, or for which the subject disagreed with the experimenters' reasoning were excluded from the analysis. Thus, the data reflect only the subject's ability to retrieve the facts necessary to verify a particular inference, not his or her ability to perform the necessary reasoning on those facts.

Table 4.4 shows the corrected percent correct responses in each of the four experimental conditions (identical versus paraphrase wordings crossed with 0- versus 30-minute retention interval). At both retention intervals, subjects verified inferences more accurately in the identical-wording condition than in the paraphrase condition, $F(1, 24) = 5.72$, $p < .025$. These results support the predictions outlined above. Presumably, identical wordings increased the probability that the two facts underlying an inference would be integrated in memory. Integration, in turn, enabled either of the two cues implicit in the inference (e.g., "Albert Profiro hated someone" and "King Egbert" in the inference "Albert Profiro hated King Egbert") to cue retrieval of both facts. When the facts underlying an inference were unintegrated, each of the facts had to be retrieved independently, given a single cue. Retrieval of a pair of unintegrated facts required more processing and was less likely to succeed than retrieval of an integrated representation of a pair of facts. Thus, integration of related facts facilitated performance on the inference test by facilitating retrieval of the facts necessary to verify inferences.

Table 4.4

CORRECT INFERENCE VERIFICATIONS
(percent)

	Wording of Common Information in Related Facts	
Retention Interval	Identical	Paraphrase
0 minutes	.78	.57
30 minutes	.78	.65

GENERAL DISCUSSION

The present research complements previous research on knowledge integration in two ways. First, it replicates the basic integration phenomenon under previously untested conditions. Subjects encountered related facts in the context of relatively long, meaningful texts. In addition, related facts occurred either in a single text or in two separate texts. The results reported indicate that integration can occur in both cases. Second, the research identified factors influencing whether or not subjects successfully integrate a given pair of related facts. The results indicated that temporal proximity and similarity of wording facilitate integration. These effects follow directly from the few simple assumptions regarding basic memory structures and processes outlined above.

It is appropriate, at this point, to introduce plausible alternative accounts of integration. Like previous researchers, we have assumed that integration is a structural phenomenon that occurs during storage of related facts. Let us consider the alternative view that individually presented facts always generate independent memory representations and that integration is a retrieval phenomenon. In other words, assume that subjects exhibit knowledge of higher-order ideas or inferences by retrieving and appropriately combining the necessary independently stored constituents at test time.

By adopting some of the same assumptions as in the structural model, the retrieval model can account for most of the results reported. Assume again that memory representations preserve lexical information and that direct activation is more reliable than associative activation. The retrieval hypothesis simply assumes that simultaneous activation of related facts must occur at test time, rather than at storage time. The dynamics of this simultaneous activation remain the same. The retrieval model accounts for the observed effects of the wording manipulation by assuming that identical wordings of related facts permit them to cue one another directly. Paraphrase wordings require associative activation via semantic mediators. Therefore, integration is more likely to succeed in the former case than in the latter.

The retrieval model accounts for the effect of presenting related facts consecutively in one story by assuming that facts presented in close temporal proximity have proximate memory locations. Retrieval of a related pair of facts can be accomplished either by retrieving the two facts directly or by retrieving one and searching for the second. If the second fact is stored near the first, it is more likely to be retrieved than if it is stored further away. Therefore, integration is more likely to succeed if related facts have occurred consecutively than if they have occurred in two different stories.

The retrieval model can also account for the interaction between wording and temporal proximity of related facts. It incorporates a simple variant of the structural model's assumption: Proximate memory locations for related facts facilitate associative as well as direct activation.

The only result that challenges the retrieval model occurred in Experiment 6. Subjects responded "OLD" to substantial numbers of NEW test items that integrate OLD test items. The model has no obvious mechanism for handling this result. Because the result has been replicated many times, by many different investigators, it is a serious problem for the retrieval model.

We must consider a second, alternative explanation for the results of Experiment 7. Subjects might actually draw and store the inferences themselves while studying the source texts. These, rather than integrated representations of the underlying facts, might be the basis for correct inference verification on the subsequent test. Again, by adopting assumptions similar to those of the structural model, the literal inference model can account for the wording effects observed in Experiment 7. However, it can be criticized on "common sense" grounds. Because subjects did not know in advance which inferences would be tested, they would have to have stored all of the many possible inferences based on facts in the two stories. It is unlikely that subjects could have done so or that people generally draw and store all possible inferences from the information they acquire.

While both of the alternative models described above have problems, they can account for many of the results reported. Of course, we cannot attribute the same significance to these post hoc explanations that we attribute to the structural model's prediction of the results. By the same token, however, we cannot rule them out. It would be too extreme to conclude from these observations that integration is invariably a strictly structural phenomenon. It seems obvious that people sometimes draw two distinct memories together specifically to evaluate a hypothetical inference. It also seems obvious that people sometimes generate and store inferences from source material they read. While structural integration probably underlies many of the observed integration phenomena, it is likely that both of the alternative processes described above also occur occasionally. Additional research is needed to elucidate the conditions under which each integration process occurs.

Regardless of which model one adopts, the present results imply certain theoretical constraints. In these experiments, both wordings and temporal proximity of related facts influenced whether or not subjects integrated the facts. Any viable model of integration must include assumptions to account for these effects.

V. CONFIGURAL EFFECTS IN HUMAN MEMORY

Chapter IV investigated two factors influencing the integration of related facts into a single, coherent structure. Frequently, such facts can serve as premises from which logical inferences can be made. For example, consider the following two facts:

(1) Albert Profiro hated all dictators.
(2) King Egbert was a dictator.

These two facts can be configured logically as

(1-2) Albert Profiro hated all dictators, one of whom was King Egbert.

to provide a basis for the inference:

(3) Albert Profiro hated King Egbert.

Given (3) as a hypothetical inference, an effective reasoner should be able to reason backward to verify (3) on the basis of (1) and (2). To do this, the reasoner would have to identify (1) and (2) as being relevant to (3), configure them appropriately, and perform the deduction.

This chapter shows that identifying and configuring facts in order to support hypothesized inferences is extremely difficult unless the facts have been committed to memory. Simply reading relevant texts for familiarization and then referring to them as needed provides an inadequate basis for deductive logic. Further, apprehension of the logical configuration of premises underlying a particular inference can be an essentially automatic process for the reasoner who has structured the facts appropriately in memory.

8

The research reported in this chapter began with an unexpected result, obtained from an unpublished experiment conducted in our laboratory. Subjects were given texts containing pairs of facts such as (1) and (2). They were then asked to verify inferences such as (3). Distractors, whose truth values were indeterminate, were also included in the inference test. In a TEXT condition, subjects simply glanced over the texts in order to become familiar with the subject matter. During the inference test, these subjects were permitted to refer back to the texts. In a MEMORY condition, subjects were instructed to study the texts carefully, attempting to learn as much of the information they contained as possible. During the inference test, these subjects were not permitted to refer back to the texts.

The most intuitively obvious prediction for the outcome of this experiment is that subjects should have performed better in the TEXT condition than in the MEMORY condition. The texts were short enough (approximately 200 words) to be read in a few minutes. It would seem that given freedom to refer back to the texts during the inference test, subjects in the TEXT condition should have been able to find the facts necessary for inference verification and therefore should have performed very well. The texts were too long to have been committed perfectly to memory, so subjects in the MEMORY condition must have forgotten many of the facts necessary to verify inferences before the inference test was given. Thus, they could not be expected to have performed as well as subjects in the TEXT condition.

The results of the experiment were straightforward: Subjects performed comparably well in both conditions. Subjects in the TEXT condition responded correctly to 81 percent of the true inferences, compared to 85 percent for subjects in the MEMORY condition. Subjects in the TEXT condition responded correctly to 79 percent of the false inferences, compared to 85 percent for subjects in the MEMORY condition. Neither difference is significant, and both differences actually favor performance in the MEMORY condition.

We believe that the explanation for this finding can be summarized in two assertions: (1) People do not know how to search an external information source, such as a text, for configural information. That is, they do not know how to search for logical configurations of facts that support hypothesized inferences. (2) Automatic memory mechanisms frequently organize acquired information so that configural information is directly accessible. That is, related facts are frequently stored together in memory in a meaningful configuration and are directly accessible as an integrated data structure. In the next section, we elaborate this view of the search and memory mechanisms underlying inference verification.

MODELING INTEGRATIVE INFERENTIAL REASONING

To verify an inference in the present experiments, a subject must perform a "backward" reasoning task. That is, he must search the texts or memory for two critical facts which, simultaneously considered and appropriately configured, permit logically valid deduction of the test inference. Consider the example introduced above:

(3) Albert Profiro hated King Egbert.

Any number of pairs of possible facts might justify (3), including, for example, the following:

(1) Albert Profiro hated all dictators.
(2) King Egbert was a dictator.

(4) Albert Profiro was a radical.
(5) All radicals hated King Egbert.

(6) Albert Profiro knew about King Egbert.
(7) Everyone who knew about King Egbert hated him.

We refer to the pair of facts that justify an inference as "critical facts." Thus, the subject begins the search for critical facts without knowing exactly what those facts might be. The only guidance the subject has is that each fact must refer to at least one of the concepts or actions involved in the inference. (Note that this is not always a simple matter of keyword referencing. The conceptual relationships between critical facts and the associated inference may be masked by synonym substitution, extensive paraphrasing, or specialization-generalization relations. These complications did not occur in the present experiments.) Further, since the texts in this experiment are cohesive stories, the concepts and actions in the inference occur in many different facts. Thus, it is necessary for the subject to search a relatively large set of facts, including many "candidate" critical facts (those referring to critical concepts or actions), for an unknown pair of facts that could be configured in some way to justify the inference logically.

Ideal Versus Actual Search Mechanisms

Let us consider performance of the backward inference task in the TEXT condition. While the task is a difficult one, the way that a simple-minded computer program could perform it is obvious. First, the program would examine the texts systematically for facts referring to the critical concepts or actions. The set of candidate facts identified must contain any pair of critical facts which, when appropriately configured, constitute a valid proof. Next, the program would formulate every possible pair of candidate facts, attempting to configure each pair so as to justify the inference. In the course of this activity, it would either encounter a pair of critical facts that logically justified the test inference and respond "TRUE," or it would exhaust the set of pairs and respond "FALSE."

The computer program described above illustrates the kind of search people "ought" to perform. Potentially relevant candidate facts might have occurred anywhere in the source texts. Only a systematic and exhaustive search of the source information, regardless of whether that information resided in memory or in an external text, would guarantee detection of all candidate facts. Similarly, any two candidate facts, when properly configured, might support the test inference. Only systematic formulation of all possible pairs of candidate facts would guarantee detection of the pair of critical facts underlying a true test inference. Systematic formulation of all pairs of candidate facts is also necessary because of the characteristics of human memory. Efforts to logically configure candidate facts presumably occur in "working memory." Given the severe limitations on human working memory capacity (cf. Miller, 1956), only a systematic pairwise consideration of candidate facts would guarantee simultaneous residence of the two critical facts in working memory. While a variety of algorithms exist that could accomplish a systematic search, it is obvious that our subjects did not use any of them. If they had, they, like the hypothetical computer program, would have performed perfectly on the inference test in the TEXT condition.

Hypothetical Memory Mechanisms

Now let us consider performance of the backward inference task in the MEMORY condition. Return to the example discussed above. In order to verify the inference

(3) Albert Profiro hated King Egbert.

the subject must retrieve two critical facts from memory:

(1) Albert Profiro hated all dictators.
(2) King Egbert was a dictator.

As we have noted, the only cues available to the subject are the concepts and actions involved in the inference. Thus, the subject eventually must cue retrieval of the first fact with the information "Albert Profiro hated ..." and cue retrieval of the second fact with the concept "King Egbert." In addition, the subject presumably must retrieve the two facts in close temporal proximity so that they can reside simultaneously in working memory. How can this happen?

We assume that, once learned, pairs of critical facts are likely to have been stored in integrated memory representations, as illustrated in Fig. 5.1. Each of the critical facts is represented as a configuration of nodes representing concepts, properties, actions, etc. The relations among the concepts, properties, and actions expressed in the fact are represented as associative connections among the nodes. The representation in Fig. 5.1 is integrated because the two fact representations share a common subrepresentation of the commmon concept, "dictator." (See Hayes-Roth and Thorndyke (1977) for a discussion of some of the factors determining whether related facts are stored in integrated or independent memory representations.)

Given an integrated memory representation of the two critical facts, the following search processes are postulated. Each of the concepts and actions in the test inference automatically retrieves memory representations of all facts that refer to it. Because the two critical facts are integrated in memory, retrieving either one of them entails associatively retrieving the other. Thus, the integrated memory representation can be retrieved by either of the two independent cues available in the inference. (Hayes-Roth and Thorndyke (1977) proposed this view of the structures and processes underlying knowledge integration. Anderson and Bower (1973), Hayes-Roth (1977), Hayes-Roth and Hayes-Roth (1975), Potts (1977), and Rumelhart, Lindsay, and Norman (1972), among others, have advanced similar views.)

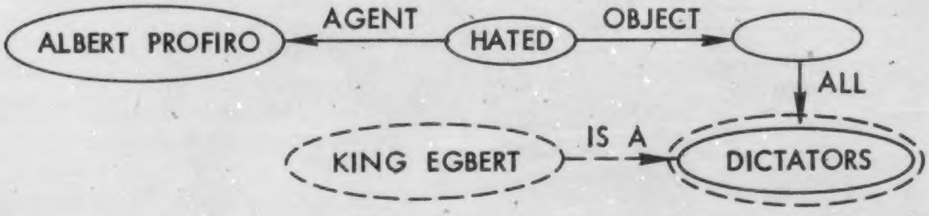

Fig. 5.1--Integrated representations of facts (1) and (2)

Because two independent cues "collaborate" to retrieve the integrated memory representation, its retrieval is more likely and more rapid than retrieval of any other cued facts. That is, the integrated representation of the critical facts is retrieved and available for subsequent processing before any non-critical candidate facts are retrieved. Thus, other candidate facts are unlikely to distract the subject from the critical facts or to interfere with their simultaneous residence in working memory.

Once retrieved, the integrated representation provides a ready basis on which to perform the necessary inferential logic. As illustrated in Fig. 5.1, the hypothetical inference (3) is implicit in the structure of the integrated memory representation of the two critical facts. Thus, searching for relevant facts and reasoning across separately acquired facts to verify a test inference reduce to essentially a recognition process. Information in a true test inference automatically cues retrieval of an integrated memory representation of the two critical facts in which the inference itself is implicit and apparent.

In order to test these assumptions, we replicated the experiment described above, with several modifications. The most important modification was the requirement that subjects provide verbal protocols of their thoughts while attempting to verify test inferences. In particular, subjects were asked to indicate how they knew that a test inference was true or false. This procedure permitted precise determination of whether a correct response followed from correct reasoning from the appropriate facts or from some other (invalid) process. The protocols also provided a detailed record of the decision processes underlying subjects' judgments of individual test inferences. Thus, analysis of the protocols provided a basis for modeling the observed differences in performance in the TEXT and MEMORY conditions.

EXPERIMENT 8

Materials. Four pairs of texts concerning the mythical country Morinthia were used (See Chap. IV). Each individual text was about a different topic, but the texts within a pair were about related topics. The four pairs of topics were (1) The First Morinthian Revolution, and The Imprisonment of Albert Profiro; (2) Religious Customs and Beliefs in Morinthia, and Religious Overtones During the Fever Epidemic in Morinthia; (3) The Marriage of Princess Isadora, Successor to the Throne of Morinthia, and The Romance Between Princess Mathilde and Basil; and (4) The Provincial Lifestyle in Morinthia, and The Morinthian Home of the Caledian Ambassador.

Each pair of texts included four pairs of critical facts, such as (1) and (2) above. Each pair of critical facts contained the information necessary to support a particular true inference not explicitly stated in either text (such as (3) above). One set of texts (also used in Experiment 7) was shown earlier, on p. .

Procedure. Subjects were tested individually. Each subject was given a booklet containing the experimental texts and the inference tests. Subjects' progress through the booklets was self-paced. Subjects were informed of the nature of the inference test they would be given.

Subjects studied each pair of texts as follows. Those in the TEXT condition simply scanned the two texts. Those in the MEMORY condition read the two texts carefully, attempting to learn as much as possible. After studying each text, these subjects completed fill-in-the-blanks tests of the information they had read. Then they referred back to the texts and corrected any errors they had made. The purpose of the tests was simply to maximize the probability that subjects committed individual facts from the texts to memory. Therefore, the test items tested memory for arbitrary information from individual sentences in the texts. They did not test any inferential knowledge, nor did they test information necessary for verifying inferences on the subsequent inference test.

After studying a pair of texts, subjects verified (judged true or false) true and false inferences derived from the texts. True inferences were defined as those that could be proved true, given the information in the texts, as illustrated by (3) above. False inferences were defined as those that could not be proved true, given the information in the texts, as illustrated by the following example.

(4) The Morinthian people were thrown into a frenzy by Albert Profiro's imprisonment.

There were four true and four false inferences, ordered randomly, on the test following each pair of texts. Subjects in the TEXT condition were encouraged to consult the studied texts during the inference test in order to determine the validity of or verify their judgments of test inferences. Subjects in the MEMORY condition were not permitted to look back at the texts. All subjects were urged to produce a continuous verbal description of their thoughts while performing the inference test. These protocols were tape-recorded and subsequently transcribed. This study-test procedure was repeated for each of the four pairs of topically related texts.

After subjects had studied and been tested on all texts, they were given a free recall test for the first text in each of the pairs of texts they had studied. Subjects were given the title of each text and instructed to write down everything they could remember from the text.

Finally, subjects were given a test of the inferences in syllogism form. On this test, each inference was immediately preceded by the two facts that presumably determined its validity. True inferences were preceded by the critical facts that supported them, as illustrated by the sequence (1), (2), (3). False inferences were preceded by facts taken from the texts that were related to the inferences but did not validate them, as illustrated by the following sequence:

The Morinthian people were thrown into a frenzy by the imprisonment of the peasant who poached the chickens.

Albert Profiro was thrown into prison.

(4) The Morinthian people were thrown into a frenzy by Albert
Profiro's imprisonment.

Subjects simply indicated whether or not each inference followed logically from the two facts that preceded it.

Design. A within-subject design was used. The one independent variable manipulated was study-test condition (TEXT versus MEMORY). All subjects worked with all four pairs of texts described above. Half the subjects worked in the TEXT condition for the first two pairs and in the MEMORY condition for the second two pairs, while the other half worked in the reverse order. In addition, within each of these two groups, each pair of texts occurred in each serial position for one of the subjects. Free recall of the first text in each pair occurred in the order of original presentation of the pairs of texts.

Subjects. Seven UCLA undergraduates and one Santa Monica College undergraduate served as subjects. Subjects were paid $6.00 for the two-hour session. In addition, subjects were given a $0.15 bonus for each correct answer above 50 percent.

Results and Discussion

Table 5.1 shows proportions of correct judgments of true and false test inferences in the MEMORY and TEXT conditions. These data represent only those test inferences to which subjects responded correctly on the final syllogism test. That is, inferences for which a subject could not perform the necessary reasoning correctly or for which the subject disagreed with the experimenters' reasoning were excluded from the analysis. (Accuracy on the syllogisms was above .98 and indistinguishable in the two conditions.) Thus, the observed differences in performance in the two conditions reflect only differences in subjects' ability to retrieve the critical facts necessary to verify inferences, not differences in subjects' ability to perform the necessary reasoning on those facts.

Table 5.1

ACCURACY ON TRUE AND FALSE INFERENCES

Measure	Condition MEMORY	 TEXT
Correct responses to true inferences	.82	.62
Correct responses to false inferences	.82	.87
Correct justifications of true inferences	.75	.48

The first two rows in Table 5.1 give proportions correct for true and false inferences. The third row gives proportions of correctly justified true inferences. These proportions are explained and discussed below. All confidence levels reported in this and subsequent sections are based on two-tailed, matched-pairs _t_-tests.

Considering first the simple proportions correct for true and false inferences, the results of the present experiment go beyond mere replication of the results discussed above. Again, no difference was observed between proportions correct on false inferences in the MEMORY and TEXT conditions (.82 versus .87), t(7)=.45. However, the proportion correct on true inferences was significantly higher in the MEMORY condition (.82) than in the TEXT condition (.62), t(7)=2.43, p <.05. Thus, the experiment replicates the observation that the availability of reference texts during inference verification does not improve performance over the case in which the information in the text must be referenced from memory. It also indicates that in some cases, exactly the opposite occurs: Storing and referencing the information from texts in memory can improve performance over that obtained when the texts themselves are available for reference.

The third row in Table 5.1 gives the proportions of true inferences judged to be true and correctly justified. Correct justification was operationalized using both strict and lenient criteria, as follows. Under the strict criterion, a true judgment was considered correctly justified only if both of the critical facts underlying the inference were explicitly stated in the protocol and linked together with appropriate logic, as illustrated in the following example:

> Inference: Anastacia DeVille was an enemy of the king.
> Fact 1: It was in prison that Albert Profiro met Anastacia DeVille.
> Fact 2: The Morinthian prison was populated exclusively by enemies of the king.

> Protocol:
> Well, that would be true because I remember reading in the other story that everybody that was in prison was an enemy of the king, and he met her when he was in prison, so she was an enemy of the king. So that's true.

Under the lenient criterion, a true judgment was considered correctly justified if at least one of the critical facts was stated in the protocol and the other was implied in what appeared to be a logically correct account, as illustrated in the following example:

> Inference: The Curfew Episode was a failure.
> Fact 1: The Curfew Episode was a revolution.
> Fact 2: All Morinthian revolutions were doomed to failure.

> Protocol:
> Well, I would say that's true because I remember reading at the end of the first story that the Spring Episode was a failure as was [sic] all the other revolts. So that kind of predestined the fact that every, every one of those things would be a failure.

> Implicit Critical Fact: The Curfew Episode was a revolution.

Two independent judges were in perfect agreement regarding the correctness of justifications according to both criteria. The lenient criterion produced a slightly smaller difference between accuracy rates in the MEMORY and TEXT conditions, but the pattern of results was the same under both criteria. In addition, both judges felt that the lenient criterion provided a more accurate evaluation of subjects' justifications. Therefore, the lenient criterion was arbitrarily adopted and the data reported in Table 5.1 reflect measurements under that criterion. (Judgments of false inferences were not evaluated for correctness of justification because of the difficulty of operationalizing correct justification for them.)

The proportion of correctly justified true inferences is a more precise measure of subjects' knowledge than simple proportion correct because it excludes both false alarms and imperfectly reasoned responses. The proportion of correctly justified true inferences was significantly higher in the MEMORY condition (.75) than in the TEXT condition (.48), $t(7) = 5.65$, $p < .01$. These data indicate that subjects' superior performance in the MEMORY condition is a consequence of their superior ability to identify the critical facts underlying true inferences.

These data support our two basic assumptions. Subjects' use of ineffective search procedures presumably impedes detection of the critical facts underlying true inferences. Configural memory mechanisms organize many learned facts, presumably facilitating detection of the critical facts underlying true test inferences. Because subjects had to rely primarily on ineffective search procedures in the TEXT condition, they should incorrectly judge many true inferences to be false. In the MEMORY condition, on the other hand, subjects should benefit from configural memory mechanisms, correctly judging most true inferences to be true. Performance on false inferences should not be affected. The results described above follow directly: While subjects performed comparably well on false inferences in the two conditions, they performed better on true inferences in the MEMORY condition than in the TEXT condition.

Protocol Analyses

As discussed above, subjects provided verbal protocols of their decision processes during inference verification. We analyzed these protocols in order to get a better understanding of the decision processes and to provide more detailed support for the proposed search and memory mechanisms. This analysis revealed six distinct types of decision processes. "Automatic retrieval of both critical facts" and "search followed by automatic retrieval" were based on the configural memory mechanisms described above. "Successful fact search" and "fact search failure" were based on heuristic search procedures. "Inference search failure" was based on an ineffective search procedure. "Faulty reasoning," as its name suggests, reflected logical errors. Table 5.2 summarizes the six decision processes and the judgments they support. We characterize each decision process briefly below.

Table 5.2

JUDGMENTS OF TRUE AND FALSE INFERENCES
PRODUCED BY SIX DECISION PROCESSES

	Validity of Test Inference	
Decision Process	True	False
Automatic retrieval of both critical facts	TRUE	...
Search followed by automatic retrieval	TRUE	...
Successful fact search	TRUE	...
Fact search failure	FALSE	FALSE
Inference search failure	FALSE	FALSE
Faulty reasoning	TRUE/FALSE	TRUE/FALSE

Automatic Retrieval of Both Critical Facts. True judgments were considered to have been based on automatic retrieval of an integrated memory representation of the two critical facts underlying the test inference if the subject verbalized only those critical facts, as illustrated in the following protocol:

True Inference: Albert Profiro hated King Egbert.
Fact 1: Albert Profiro hated all dictators.
Fact 2: King Egbert was a dictator.

Protocol:
True, because King Egbert was a dictator and Albert Profiro hated all dictators. So it's true.

We assume that this decision process involves no conscious search for relevant facts. There is no indication that the subject considered any other related facts in the attempt to verify the test inference. Rather, the subject appears to have been able to retrieve the two critical facts automatically on presentation of the test inference.

Search Followed by Automatic Retrieval. True judgments were considered to have been based on a combination of search and automatic retrieval if (a) the subject verbalized one or more facts that were related to the test inference before verbalizing the first of the two critical facts underlying the test inference, and (b) the subject did not verbalize any other facts between verbalizations of the two critical facts. This decision process is illustrated in the following protocol:

True Inference: The Caledian Ambassador's house was surrounded by lilies.
Fact 1: The Caledian Ambassador's house was surrounded by those flowers that were prevalent in Morinthia.
Fact 2: A species of lily was the prevalent flower in Morinthia.

Protocol:
OK, they said he had fences and ... um ... in Morinthia

> they have fences and they put plants or flowers around, in their, no, vines ... and they're twined through their fences. Um, ... he had fences. OK, the cottage had fences and lilies aren't vines so they can't intertwine through his fence. He had, I remember they said he had typical vegetation of Morinthia. And since lilies were one of their best flowers, the ones they liked and he had typical vegetation, then I say the statement is true. He grew lilies.

We assume that this decision process involves an initial search of either the texts or memory for relevant information. In the protocol above, the search appears to have been for relevant facts, although, in other cases, it appears to have been for the test inference itself. In either case, the initial search leads to detection of one of the two critical facts necessary to verify the test inference. That fact then cues automatic retrieval of the integrated memory representation of both critical facts.

Successful Fact Search. True judgments were considered to have been based on successful fact search if the subject verbalized one or more facts that were related to the test inference before verbalizing each of the two critical facts, as illustrated in the following protocol:

True Inference: There were berries growing outside of the Caledian Ambassador's cottage.

Fact 1: The native shrub of Morinthia hugged the walls of the Caledian Ambassador's cottage.

Fact 2: There was also a native shrub, called the salsa shrub, that had broad leaves and was covered with red berries.

Protocol:

> OK, that's the second story. Be sure to look there

> because there's nothing in the first story about the Caledian Ambassador. So, um, [reading] it talks about the cottage first and, uh, lush shrubs hug the walls of the cottage. OK, it was brought out about flowers that were prevalent in Morinthia. OK, I don't know if those flowers had berries or not. I look at the first story to see that. It doesn't say anything about berries in the second story. OK, um, first story ... They talk about the people. They don't talk about Ambassadors, although I'm not sure that Ambassadors are really that much different than people. OK, a species of lily. Uh, I don't see anything about lilies in the second story. I'm not sure that's relevant. OK, there's a native shrub, also quite popular, called the salsa shrub. Its leaves were broad and it had berries. Um, if that's the same shrub as the shrub in the second story, I would say that yes, there were berries.

Successful fact search reflects active search of either the texts or memory for both critical facts underlying a true test inference. These searches indicate an awareness that two critical facts are needed and that they might occur in either of the two relevant texts. Thus, successful fact search reflects a heuristic, rather than algorithmic, search procedure. In addition, we assume that the occurrence of successful fact search implies that the critical facts underlying the true test inference have not been stored in an integrated memory representation. Indeed, it is the only way subjects can correctly verify true inferences whose critical facts have not been integrated in memory.

Fact Search Failure. False judgments were considered to have been based on fact search failure if the subject verbalized a number of alternative facts before concluding that the test inference was false. A false judgment of a true test inference based on fact search failure is illustrated in the following protocol:

True Inference: The Caledian Ambassador's cottage was surrounded by lilies.

Fact 1: The Caledian Ambassador's cottage was surrounded by those flowers that were prevalent in Morinthia.

Fact 2: The most prevalent flower in Morinthia was a species of lily.

Protocol:

Uh, the second story. I'll look again. Probably should remember this by now. Um, [reading] lush shrubs, again, nothing about lilies. He had a low white fence and there's something in the first story about fences having shrubs over them. Uh, looking at that now. OK, the most prevalent flower in Morinthia was a species of lily. OK, um, I see nothing about ... no necessary implication between fences and gates which have vines on them and lilies. So I'd say that, uh, the Caledian Ambassador's cottage was surrounded by lilies is false.

Fact search failure reflects active search of either the texts or memory for facts that are related to a test inference. These searches also indicate an awareness that two critical facts are needed and that they might occur in either of the two relevant texts. Thus, fact search failure also reflects a heuristic search procedure.

Inference Search Failure. False judgments were considered to have been based on inference search failure if the subject explicitly searched for the test inference, itself. A false judgment of a false test inference based on inference search failure is illustrated in the following protocol:

False Inference: The people threw stones during the first Morinthian revolution.

Protocol:

I'm going back to the first one because I remember in the second one they threw stones in the street, but I'm not sure about the first one. So, starting in the middle of the page, [reading] "It provided the ... Everyone who disobeyed the law was punished severely. The peasant was branded.... The townspeople were thrown into a frenzy with the severity of the sentence. The crowd stormed the palace " No, they didn't throw rocks. It didn't say on here.

Inference search failure is based on an ineffective search procedure. The subject assumes that if the test inference were true, it would appear verbatim in a text or in memory. Although the subject verbalizes several facts from the text, these facts are selected not because they might be useful in proving the validity of the test inference, but rather because the subject simply happened to encounter them while searching for the literal inference.

Faulty Reasoning. True and false judgments were classified as faulty reasoning if there was a logical error in the subject's justification of a judgment of a test inference, as illustrated in the following protocol:

True Inference: Albert Profiro hated King Egbert.
Fact 1: Albert Profiro hated all dictators.
Fact 2: King Egbert was a dictator.

Protocol:

OK. In the second sheet, the Curfew Episode, they mention the name Albert Profiro and he, um, he wanted to ... uh ... he wanted to talk to the people to tell them, um, to return to their homes because of the Curfew law. Um, since Egbert was the ruler, yeah, was the ruler of Morinthia, um, he was really strict like I said before;

> that is, something small, a small crime or something happens, then they get punished severely. So, um, the soldiers mistook Albert of starting a riot instead of telling the people to go home because it's a curfew and so he was thrown into prison. Therefore, I think Albert hated King Egbert because of the misunderstanding and how bad he had to suffer for it.

Implicit Unsupported Premise: Albert Profiro would hold King Egbert responsible for the misunderstanding and hate him for it.

We analyzed subjects' response protocols to determine how often each of the six decision processes occurred. Each judgment of a test inference was categorized as exemplifying one of the decision processes according to the rules described above. Two independent judges were in complete agreement regarding the categorization of subjects' judgments. Table 5.3 summarizes the results of the protocol analyses. Each entry in Table 5.3 gives the mean proportion of responses to each item type that exemplified each of the decision processes. Thus, each column in the table sums to 1.0.

Table 5.3

PROPORTIONS OF JUDGMENTS BASED ON EACH TYPE OF DECISION PROCESS

	Condition			
	MEMORY		TEXT	
	Validity of Test Inference			
Decision Process	True	False	True	False
Automatic retrieval of both critical facts	.70	...	.11	...
Search followed by automatic retrieval	.05	...	.28	...
Successful fact search	.00	...	.09	...
Fact search failure	.05	.46	.04	.25
Inference search failure	.12	.33	.30	.55
Faulty reasoning	.09	.21	.19	.20

In the MEMORY condition, 70 percent of all judgments of true inferences were correctly justified true judgments based on automatic retrieval of both critical facts. The remaining judgments of true inferences were fairly evenly distributed among the remaining categories of decision processes. Correct judgments of false inferences were primarily based on fact search failure (.46). However, there were also large numbers of judgments based on both inference search failure and faulty reasoning.

In the TEXT condition, most judgments of true inferences were either "true" judgments based on search followed by automatic retrieval (.28) or "false" judgments based on inference search failure (.30). The remaining judgments of true inferences were fairly evenly distributed among the remaining categories of decision processes. Correct judgments of false inferences were primarily based on inference search failure (.55). However, there were also large numbers of judgments based on both fact search failure and faulty reasoning.

The overall pattern of results shown in Table 5.3 is consistent with the proposed account of performance of the inference verification task. In the MEMORY condition, subjects performed well on the true inferences because they had good memory for the critical facts underlying the inferences. In the TEXT condition, subjects performed poorly on the true inferences because they had poor memory for the critical facts. In neither condition did subjects perform well on true inferences by applying effective search procedures. Similarly, subjects performed reasonably well on false inferences only because the ineffective search procedures they applied in both conditions happened to lead to the correct judgment (false). These conclusions are supported by the following separate analyses of decision processes underlying correctly justified true judgments and those underlying false judgments and incorrectly justified true judgments.

Decision Processes Underlying Correctly Justified True Judgments. Table 5.4 shows the conditional proportions of correctly justified true judgments based on each of the three decision processes that produced them. In the MEMORY condition, 93 percent of all correctly justified true judgments were based on automatic retrieval of both critical facts underlying the inferences. In fact, for six of the eight subjects, all correct justifications of true inferences were based on this decision process. Only two of the eight subjects engaged in any search activity at all during correct justifications of true inferences in the MEMORY condition. An additional 7 percent of correctly justified true judgments in the MEMORY condition were based on search followed by automatic retrieval. Summing the two, all correctly justified true inferences in the MEMORY condition can be attributed to configural memory mechanisms. In the TEXT condition, only 28 percent of all correctly justified true judgments were based on automatic retrieval of both critical facts. However, an additional 58 percent were based on search followed by automatic retrieval. Summing the two, 86 percent of all correctly justified true inferences in the TEXT condition can be attributed to configural memory mechanisms. Only three of the eight subjects performed any successful searches at all. The difference between the proportions of correctly justified true judgments attributable to configural memory mechanisms in the two conditions is not significant, $t(7)=1.82$, $p > .1$.

Table 5.4

PROPORTIONS OF CORRECTLY JUSTIFIED TRUE JUDGMENTS BASED ON EACH TYPE OF DECISION PROCESS

	Condition	
Decision Process	MEMORY	TEXT
Automatic retrieval of both critical facts	.93	.28
Search followed by automatic retrieval	.07	.58
Successful fact search	.00	.14

These results support our claim that subjects' correctly justified true judgments derive from automatically cued memory for the critical facts underlying the test inferences, rather than from search procedures that detect the critical facts. Correctly justified true judgments in both conditions were produced almost exclusively by decision processes based on configural memory mechanisms. Thus, as we have suggested, subjects' excellent performance in the MEMORY condition reflected their having learned the critical facts underlying test inferences and, consequently, their ability to exploit configural memory mechanisms. In the TEXT condition, on the other hand, subjects made no effort to learn the information in the texts they studied. Therefore, their knowledge of the critical facts underlying test inferences and their ability to exploit configural memory mechanisms were critically dependent upon incidental learning. Interestingly, virtually all (86 percent) of the correctly justified true judgments in the TEXT condition were based on incidental learning of the critical facts. Apparently, incidental learning was reasonably high as a consequence of subjects having searched each pair of source texts eight times (four times for true test inferences and four times for false test inferences).

Analysis of the temporal intervals separating particular verbalizations during inference verification provides additional support for this position. The proposed model assumes that automatic retrieval of both critical facts involves no search activity. Retrieval of the two critical facts presumably occurs simultaneously on presentation of the test inference. Therefore, the temporal intervals preceding verbalizations of the two critical facts should not include any search time. That is, the temporal interval separating "offset" of the test inference from "onset" of verbalization of the first critical fact and the temporal interval separating offset of verbalization of the first critical fact from onset of verbalization of the second critical fact should not include any search time.

The model does not predict that these "verbalization times" should be zero, because factors other than search activity influence verbalization times. For example, determination that the information retrieved constitutes justification of the test inference, formulation of a verbal description, and initiation of verbalization should take a certain amount of time. Even a binary OLD/NEW recognition judgment of simple sentences can take on the order of 1.5 seconds (cf. Anderson, 1974). Inference verification judgments are considerably more complex than recognition judgments. In addition, verbal justifications are considerably more complex than the simple button-pressing response typically required in recognition paradigms. Therefore, we would expect inference verification to take considerably longer than the 1.5 seconds required for sentence recognition.

Another factor influencing verbalization times is time constraints. Obviously, verbalization times would be shorter if subjects were constrained to respond as quickly as possible. Subjects were under no time constraints in the present experiment. Quite the contrary, they were being rewarded for accuracy only. They were free to pace themselves however they liked, and they apparently took their time. The mean response time (interval between offset of the inference and verbalization of the judgment "true") for correctly justified true judgments was 28.8 seconds in the MEMORY condition and 42.3 seconds in the TEXT condition.

For the above reasons, we arbitrarily set a criterion time at 5 seconds. Verbalization times under 5 seconds were taken to indicate automatic retrieval; those over 5 seconds were taken to indicate the occurrence of search activity. This criterion seemed to be reasonably conservative and was decided upon before any times were measured.

Table 5.5 lists the proportions of correctly justified true judgments for which verbalization times for both critical facts, only the second critical fact, or neither critical fact were less than 5 seconds. Note that the proportions in Table 5.5 correspond directly to those in Table 5.4. True judgments for which verbalization times for both critical facts were less than 5 seconds correspond to automatic retrieval of both critical facts. Those for which only verbalization times for the second critical fact were less than 5 seconds correspond to search followed by automatic retrieval. Those for which neither verbalization time was less than 5 seconds correspond to successful search.

Table 5.5

PROPORTIONS OF VERBALIZATION TIMES LESS THAN FIVE SECONDS DURING CORRECT JUSTIFICATION OF TRUE INFERENCES

	Condition	
Critical Fact	MEMORY	TEXT
Both critical facts	.73	.20
Fact 2 only	.16	.51
Neither critical fact	.11	.29

According to the 5-second criterion, 73 percent of all correctly justified true judgments in the MEMORY condition involved automatic retrieval of both critical facts. An additional 16 percent involved search followed by automatic retrieval. Summing the two, 89 percent of all correctly justified true judgments in the MEMORY condition can be attributed to configural memory mechanisms. In the TEXT condition, only 20 percent of all correctly justified true judgments involved automatic retrieval of both critical facts. However, an additional 51 percent involved search followed by automatic retrieval. Summing the two, 71 percent of all correctly justified true judgments in the TEXT condition can be attributed to configural memory mechanisms. The difference between the proportions of correctly justified true judgments attributable to configural memory mechanisms in the two conditions is not significant, $t(7)=1.48$, $p > .1$.

These results provide additional support for the claim that subjects' correctly justified true judgments derive from automatically cued memory for the critical facts underlying the test inferences, rather than from search procedures that detect the critical facts. As suggested by the protocol analyses described above, the majority of correctly justified true judgments in both conditions were produced by decision processes based on configural memory mechanisms. Note that the proportions of judgments attributable to configural memory mechanisms shown in Table 5.5 are somewhat lower than those in Table 5.4. However, the pattern of estimates is the same. In the MEMORY condition, most correctly justified true judgments involved automatic retrieval of both critical facts, while a smaller number involved search followed by automatic retrieval. In the TEXT condition, most correctly justified true judgments involved search followed by automatic retrieval, while a smaller number involved automatic retrieval of both critical facts. In both tables, the total proportion of judgments attributable to configural memory mechanisms was slightly and non-significantly higher in the MEMORY condition than in the TEXT condition. The only noteworthy difference between the results shown in Table 5.4 and those in Table 5.5 is the suggestion in Table 5.5 that a fair number of correctly justified true judgments in both conditions may have involved successful search. However, this may be a consequence of our having set the 5-second criterion too low, underestimating the time spent formulating and initiating a verbalization in the absence of speed stress.

Decision Processes Underlying False and Incorrectly Justified True Judgments. Table 5.6 shows the conditional probabilities that false judgments and incorrectly justified true judgments were based on each of the three decision processes that produce these judgments. In the MEMORY condition, 39 percent of these judgments were based on fact search failure, 36 percent were based on inference search failure, and 25 percent were based on faulty reasoning. In the TEXT condition, 18 percent of these judgments were based on fact search failure, 57 percent on inference search failure, and 25 percent on faulty reasoning. The difference between the proportions of judgments attributable to fact search failure versus inference search failure in

Table 5.6

PROPORTIONS OF FALSE AND INCORRECTLY JUSTIFIED TRUE
JUDGMENTS BASED ON EACH TYPE OF DECISION PROCESS

	Condition	
Decision Process	MEMORY	TEXT
Fact search failure	.39	.18
Inference search failure	.36	.57
Faulty reasoning	.25	.25

the two conditions is not significant, t(7)=1.23, p > .2. Obviously, faulty reasoning also occurs comparably often in the two conditions.

These results support our claim that subjects' false judgments and incorrectly justified true judgments derive primarily from ineffective search procedures and logical errors, rather than from errors during execution of an effective search algorithm. It should also be pointed out that our classification criterion for inference search failure was conservative. A false judgment was classified as inference search failure only if two judges independently agreed that the subject had given explicit indication that the false judgment was based on failure to find the test inference explicitly in the texts or memory for the texts. Thus, the proportions for inference search failure given in Table 5.6 are probably underestimated, while the proportions for fact search failure are probably overestimated. In addition, the decision process identified as fact search failure does not reflect a particularly effective systematic search procedure. This decision process was operationalized simply as consideration of more than one fact from the texts prior to making a judgment of "false." No protocol provided any evidence of a subject's having made a systematic search of the available facts, such as that performed by the hypothetical computer program, before rendering a judgment of "false." Thus, these results indicate that subjects did not use an effective search algorithm in searching either the available texts or memory of the texts for facts relevant to verifying test inferences.

Comparison of Judgment Times in MEMORY and TEXT Conditions

Table 5.7 shows mean judgment times (interval between offset of the inference and verbalization of the judgment) for correctly justified true and correct false judgments in the MEMORY and TEXT conditions. These times presumably include component times representing decision processes (search for and retrieval and evaluation of memory representations) and verbalization processes (determination, formulation, and initiation of appropriate verbalizations regarding the decision processes). Because many of these components differ for true and false judgments, it is not useful to compare response times for different judgments. However, it is

Table 5.7

MEAN JUDGMENT TIMES FOR CORRECT JUDGMENTS
(sec)

	Condition	
Judgment	MEMORY	TEXT
True	28.8	42.3
False	20.3	42.1

interesting to compare response times for each of these judgments across the two conditions.

As might be expected, based on the component times discussed above, correctly justified true judgments were made faster in the MEMORY condition than in the TEXT condition, t(7)=2.79, $p < .05$. This difference can be attributed to differences in the decision processes underlying correct justifications of true inferences in the two conditions. Subjects completed their correct justifications of true inferences faster in the MEMORY condition because those justifications were more frequently based on automatic retrieval of both critical facts and rarely involved any search activity at all. In the TEXT condition, on the other hand, even though most correct justifications of true inferences terminated with automatic retrieval of integrated memory representations, they frequently began with some kind of search activity. Thus, correctly justified true judgments took more time in the TEXT condition than in the MEMORY condition because more search activity was necessary.

Correct false judgments were also faster in the MEMORY condition than in the TEXT condition, t(7)=5.34, $p < .002$. The most reasonable explanation for this difference is that the decision processes underlying false judgments are faster in the MEMORY condition than in the TEXT condition. Subjects apparently search memory faster than they search a reference text.

Memory for Source Information

In addition to facilitating inference verification, another potential advantage of the MEMORY condition is memory for the information in the texts. Of course, subjects were encouraged to learn the information in the MEMORY condition, but not in the TEXT condition. However, they spent a considerable amount of time examining the texts, even in the TEXT condition. This might be expected to produce some memory for the information, and our observation that most of the correctly justified true judgments in the TEXT condition were attributable to configural memory mechanisms suggests that there was incidental learning in the TEXT condition.

Subjects' memory for the information in the texts was analyzed as follows. The texts were broken into propositional units (Thorndyke, 1977), then subjects' recall performance was scored for recall of the gist of each proposition in the appropriate texts. Thus, it was possible to determine the proportion of propositions recalled by each subject in each condition. The data are given in Table 5.8. (Since one subject did not have time to perform the recall task in the TEXT condition, the data are based on responses from seven subjects.)

The first row in Table 5.8 shows recall proportions for the two conditions. Recall is considerably better in the MEMORY condition (.61) than in the TEXT condition (.30), t(6)=4.68, p < .01. Thus, even though subjects inspected the texts thoroughly in the TEXT condition, they learned considerably less than in the MEMORY condition, where learning was intentional.

The second row in Table 5.8 shows the recall proportions for propositions that subjects used to correctly justify true inferences

Table 5.8

RECALL OF SOURCE TEXTS

	Condition	
	MEMORY	TEXT
All propositions	.61	.30
Target propositions	.80	.76
Non-target propositions	.58	.27

in the two conditions. The third row shows the recall proportions for propositions that subjects did not use to correctly justify true inferences. These proportions approximate the following breakdown. Propositions used to justify true inferences represent those that subjects identified as "targets" in the text or in memory and used in correct justifications of test inferences. Propositions not used to justify true inferences represent those that were simply scanned (in the TEXT condition) or unretrieved (in the MEMORY condition) during the inference test. The breakdown is approximate because other propositions must have been identified as targets for use in incorrect justifications of true and false inferences. Because of the logical errors involved in these justifications, the identities of target propositions were frequently ambiguous. Therefore, we did not attempt to include propositions from these justifications in the target group but simply classified all propositions not used in correct justifications of true inferences as non-targets. As a result, the observed differences between recall of target and non-target propositions discussed below is probably smaller than the actual difference.

In both conditions, subjects learned target propositions. Recall of target propositions was high, on the order of 80 percent, in both conditions, t(6)= .32. This suggests that identifying a proposition as the target of a search (either in memory or a text) and using it in a correct justification is a powerful learning experience. This finding is consistent with other studies (cf. Anderson & Biddle, 1975; Frase, 1975; Frase & Schwartz, 1975), indicating that subjects retain information that is directly relevant to post-test questions better than they retain other information from studied texts.

Subjects also learned many non-target propositions in the MEMORY condition, but not in the TEXT condition. Recall of non-target propositions was 58 percent in the MEMORY condition, compared to 27 percent in the TEXT condition, t(6)=4.26, p < .01. These data suggest that subjects who inspect a text with the intention of learning the information it contains retain a considerable amount of that information even after a relatively long retention interval (30 to 90 minutes) filled with highly interfering memory and inference tests. By contrast, subjects who thoroughly and repeatedly search texts for information relevant to particular inferences learn very little of the information scanned.

It might be argued that target propositions were recalled better than non-target propositions simply because they were verbalized by the subject during the inference test. However, in the TEXT condition, virtually all propositions occurred at least once and most occurred several times in subjects' protocols, yet recall of target propositions was substantially higher than recall of non-target propositions. Thus, it is not simply the verbalization of a proposition that establishes it in memory, but rather its identification as a target of the current search effort and its ultimate use in a correct justification.

Thus, the higher overall recall probability in the MEMORY condition, as compared to the TEXT condition (first row in Table 5.8), is attributable to two factors. First, subjects in the MEMORY condition identify more target propositions during the inference test and therefore learn more of them very well. This is a relatively minor factor, however, because of the relatively small number of identifiable target propositions (approximately 12 percent of all propositions in the texts). The major factor is the superior learning of non-target propositions induced by studying the texts prior to the inference test, compared to the minimal incidental learning of non-target propositions induced by searching the texts repeatedly during the inference test.

Individual Differences

The proposed model assumes that level of performance on the inference test is determined primarily by memory for the facts necessary to verify test inferences. It assumes that effective search procedures play a minor role in determining level of performance. This model accounts well for the group data discussed above. In this section, we examine the performance levels of individual subjects. There was considerable variability in the absolute levels of performance observed for individual subjects. If the proposed model is correct, it should be able to account for variation in performance level among individuals as well as variation in performance level between MEMORY and TEXT conditions. Individuals with good memories should perform better on the inference test than individuals with poor memories. The quality of individuals' search procedures should have little effect on performance on the inference test.

Table 5.9 presents summary data regarding the performance levels and memory and search processes of individual subjects. The first row in Table 5.9 records proportion of correctly justified TRUE judgments for each subject pooled over MEMORY and TEXT conditions. Subjects are ordered from left to right in Table 5.9, according to level of performance. Thus, Annie produced the highest proportion of correctly justified true judgments (.84), while Patrick and Donna produced the lowest proportions (.44).

Table 5.9

SEARCH, MEMORY, AND PERFORMANCE PROFILES FOR INDIVIDUAL SUBJECTS

	Subject							
Measure	Annie	Paul	Brian	Eva	Betty	Tina	Donna	Patrick
Correctly justified true judgments in MEMORY and TEXT conditions	.84	.72	.69	.64	.60	.59	.44	.44
Recall of non-target propositions in MEMORY condition	.78	.83	.65	.70	.69	.47	.38	.28
Successful search efforts in search condition	.50	.40	.00	.00	.33	.00	.00	.00

The second row in Table 5.9 provides a measure of the quality of individual subjects' memories. This variable is operationalized as the proportion of non-target facts recalled from the texts in the MEMORY condition. The measurement is restricted to non-target facts because, as discussed above, memory for target facts is better than for non-targets. Since subjects identified different numbers of target facts during the inference test, including these facts in the measurement would bias the analysis. The measurement is restricted to the MEMORY condition because one subject was unable to perform the recall test in the TEXT condition, due to lack of time. The proportion of non-target propositions recalled provides a direct measure of the quality of subjects' memories. Subjects who recalled higher proportions of non-target propositions have better memories than those who recalled lower proportions of those propositions.

The third row in Table 5.9 provides a measure of the quality of the search procedures employed by each subject. This variable is operationalized as the proportion of true test inferences (excluding those correctly justified on the basis of automatic memory mechanisms) that the subject correctly justified on the basis of successful search. Proportion of successful search efforts was computed only on data from the TEXT condition, to provide a direct measure of the quality of subjects' search procedures in the absence of any memory effects. Subjects who had higher proportions of successful search efforts have better search procedures than those who had lower proportions.

Table 5.9 indicates marked individual differences in the levels of performance and the qualities of the memory and search mechanisms of individual subjects. The individual differences appear to be orderly and consistent with the theoretical framework developed above. First, performance on the inference test appears to be influenced strongly by the quality of a subject's memory. There is a positive correlation between proportion recall of non-target propositions and proportion of correctly justified true judgments, $r = .89$, $p < .01$. Second, performance on the inference test appears to be only minimally dependent upon the quality of a subject's search procedures. Most subjects had no successful search efforts, and the correlation between proportion of successful search efforts and proportion of correctly justified true judgments is not significant, $r = .69$, $p < .10$. Thus, as predicted by the proposed model, a subject's performance on the inference test was determined by ability to remember information from the stories, not by ability to search effectively for particular information.

It is interesting to note that the two best subjects, Annie and Paul, had the best search procedures as well as the best memories. There is a significant positive correlation between proportion recall of non-target propositions and proportion of successful search efforts, $r = .72$, $p < .05$. These results indicate that subjects who have good memories also tend to have relatively good search procedures. Note, however, that even subjects with good search procedures were heavily dependent upon automatic memory mechanisms in their efforts to verify test inferences. Thus, the highest observed proportion of successful search efforts was only 50 percent. Further, all three of the subjects who exhibited good search procedures based the majority of their correct justifications of true test inferences (78 to 90 percent) on automatic memory mechanisms.

PRACTICAL IMPLICATIONS

Peoples' ability to detect and use logical configurations of related facts apparently derives from highly developed memory mechanisms. It is extremely difficult to detect configural information in an external information source, such as a text. It is much easier for people to detect configural information if the source information has been committed to memory. The model proposed to account for these effects assumes that (1) subjects use search procedures that are inadequate to detect configural information, and (2) automatic memory mechanisms organize acquired facts in memory structures that make configural information salient and highly accessible. In addition to predicting the details of group performance in the MEMORY and TEXT conditions, the proposed model accounts well for individual differences.

These results have obvious practical implications for anyone who deals with information and wants to perform more sophisticated information processing than simple fact retrieval: People should not rely upon the ability to search available information sources for relevant information as it is needed. Most people employ search procedures that are inadequate for that task. Instead, they should attempt to commit as much important information as possible to memory. Once information has been memorized, it is available not only for simple fact retrieval, but for more sophisticated information processing, such as deductive reasoning, as well. These considerations are particularly relevant in education.

The value of "memorization" in education is widely disputed. Those who doubt the value of memorization usually base that opinion on a preference for more ambitious educational goals. Arguments in this category usually run as follows. Students should understand, rather than memorize, the information they encounter. Students are deluged with facts that are related in various ways. These facts arrive at different times and in different contexts. Students should not simply learn the individual facts they encounter--they should integrate them. That is, students should construct a meaningful conceptual structure in which individual facts can be embedded. This is what is meant by understanding. It is the basis for the ability to perform higher-order information processing, such as comparison, characterization, and inferential reasoning. Students should learn to use facts in these important ways, rather than simply to reproduce them. Therefore, it is less important for students to memorize facts than it is to know where to find information and how to use it when it is needed.

Students frequently offer a similar argument when requesting open-book examinations. The claim is that traditional closed-book examinations emphasize the "wrong" thing, namely, memory for specific facts. In addition, closed-book examinations encourage "undesirable" study habits. Students are preoccupied with fact memorization, at the expense of efforts to integrate and understand the material. Open-book examinations, on the other hand, deemphasize memory for specific facts, since the facts are readily available in the text during the examination. Students are freed of the need to spend valuable study time memorizing facts and can, instead, devote that time to developing a better understanding of the material. The open-book examination presumably provides a good opportunity for students to demonstrate this understanding.

Let us examine these arguments in light of the present research. The inference verification task used in our experiment required subjects to understand and integrate related facts in meaningful conceptual structures, to find information when it was needed, and to use facts in a sophisticated way. Thus, ability to perform the task exemplifies the kind of educational goal advocated in the arguments above. The results of the experiment show that subjects were unable to perform the inference verification task when the necessary facts resided only in external texts (open-book test). That is, they were unable to integrate facts and understand the relationships among them. They were unable to find facts that were relevant to test inferences and therefore unable to use the facts to verify test inferences. On the other hand, subjects performed the inference verification task rather well when the necessary facts had been learned (closed-book test). They were able to integrate learned facts and understand the relationships among them. They were able to retrieve facts that were relevant to test inferences and therefore able to use learned facts to verify test inferences. Further, integration of related facts and retrieval of relevant facts during inference verification appeared to be automatic memory functions, requiring little conscious effort by the subject.

These results indicate that students are poor processors of information that resides in an external text but reasonably effective processors of information that they have learned. Consequently, attempting to perform sophisticated information processing instead of learning specific facts, as advocated in the arguments above, may be self-defeating. Apparently, learning the individual facts that are involved in a complex knowledge structure is an important, and perhaps necessary, precursor to a thorough understanding of the relationships among those facts. These results also suggest that open-book examinations may not produce their intended effects. The danger is that students will rely too heavily upon the availability of the text and fail to exploit the powerful organizational properties of human memory. Thus, students should not forgo efforts to learn course material in favor of efforts to understand and integrate it. They should work to learn course material, as well as to understand and integrate it.

In summary, the present research suggests that memorizing new information is a critical step in the learning process. We do not mean to detract from the importance attributed to understanding and sophisticated information processing as educational goals. On the contrary, we agree that simple fact acquisition is an impoverished educational goal. Good understanding of acquired information and, perhaps more importantly, the ability to apply acquired information should be the ultimate goals. However, the present results suggest that these goals are best achieved when they are predicated on a strong foundation of initial learning.

VI. TEXT ANNOTATION: A TECHNIQUE FOR FACILITATING KNOWLEDGE INTEGRATION

Chapters IV and V focused on the process of integration of information in a text. This chapter investigates a technique for improving subjects' integration of knowledge. Earlier studies of integration typically required subjects to study a text containing one or more pairs of facts from which particular inferences could be deduced (e.g., Bransford & Franks, 1971; Frase, 1969, 1973, 1975; Haviland & Clark, 1974; Hayes-Roth, 1977). Subjects were then tested on their knowledge of these inferences. For instance, Frase (1969) presented subjects with brief stories like the following:

> The Fundalas are outcasts from other tribes in Central Ugala. It is the custom in this part of the country to get rid of certain types of people. The hill people of Central Ugala are farmers. The upper highlands provide excellent soil for cultivation. The farmers of this country are peace loving, which is reflected in their artwork. The outcasts of Central Ugala are all hill people. There are about fifteen different tribes in this area. (p. 2)

Later, subjects were tested for their knowledge of inferences such as

(1) The Fundalas are hill people

which follows from

(2) The Fundalas are outcasts from other tribes in Central Ugala

and

(3) The outcasts of Central Ugala are all hill people.

Working in a similar paradigm, in Experiment 7, (Chap. IV) we presented pairs of related stories, such as those on p. 89, to subjects, who were then tested for their knowledge of the inference

(4) Albert Profiro hated King Egbert

which is based on the facts

(5) King Egbert was a dictator (from the first story)

and

(6) Albert Profiro hated all dictators (from the second story).

In both of these studies, successful integration of related facts permitted subjects to deduce information that was neither required for nor suggested by simple comprehension of the texts.

Despite the importance of information integration, people are not very good at it. Frase (1969) found that free recall of his stories included relatively few inferences, even when subjects were told explicitly to include as many inferences as possible in their recall. A related result depends on the number of inferences to which a subject is exposed. If an inference chain is represented by the linear order A -> B -> C, a subject verifing the inference "A -> C" during acquisition would be exposed to only that one inference. Subjects verifing the more complex inference "A -> E," represented by the chain A -> B -> C -> D -> E, should be exposed to the intermediate inferences of A -> C, B -> D, C -> E, A -> D, and B -> E, as well as the explicit inference A -> E. One might expect the latter condition to produce more inferences on a subsequent recall test. This did not occur.

Unpublished research from our laboratory indicates similar difficulties in information integration. Subjects (UCLA undergraduates enrolled in an introductory psychology course) read four pairs of related stories (the materials illustrated in Chap. IV). Four pairs of related facts occurred in each pair of stories, with one fact from each pair occurring in each story. Each pair of facts could be used to deduce an inference, as illustrated in (4) to (6) above. After studying all four pairs of stories, subjects received a verification test on the sixteen true inferences and sixteen distractors. Eight of the sixteen subjects performed at or below chance on this verification test (as determined by a high-threshold correction for guessing, Green and Swets, 1966). Providing three repetitions of each story did little to improve performance: Seven of the sixteen subjects still performed no better than chance. Performance on inferences was conditionalized on correct performance on a subsequent syllogism test. That is, inferences were included in the analysis only if subjects responded to them correctly when they were preceded by the appropriate facts (premises). Thus, poor performance on the inference test was not due to subjects' inability to perform the necessary reasoning, but rather to their failure to integrate the related facts from which the inferences could be deduced.

As these studies illustrate, facilitating subjects' integration of related facts and deduction of the inferences they entail is an important educational goal. The present research investigated the facilitative effects of text annotation. Subjects studied the stories described above and attempted to deduce as many inferences from them as they could. Then they were tested on their ability to verify valid inferences deduced from the pairs of related facts that had appeared in the stories.

Two types of text annotation were evaluated. In a FACT annotation condition, facts from the first story in a pair were repeated in footnotes to related facts in the second story. Thus, this condition focused the subjects' attention on pairs of related facts. In a FACT-AND-INFERENCE annotation condition, facts from the first story were again repeated in footnotes to related facts in the second story. However, in this condition, the repeated facts were accompanied by the inferences that could be deduced from the fact pairs. Thus, this condition focused subjects' attention on pairs of related facts and informed them of the appropriate inferences as well. In a CONTROL condition, texts were not annotated.

Although both annotation conditions should produce better performance than the CONTROL condition, it is not clear which of them should produce the best performance. Because the FACT-AND-INFERENCE condition provides subjects with the test inferences, they cannot fail to deduce any of them, nor can they incorrectly deduce unwarranted inferences. The FACT condition, by contrast, requires subjects to deduce inferences for themselves. Thus, subjects may fail to deduce the inferences or may incorrectly deduce unwarranted inferences. On the other hand, the kind of processing activity required by the FACT condition might provide another kind of advantage. Actively deducing an inference from two premises is a "deeper" kind of processing (Craik & Lockhard, 1972) than simply studying inferences that have already been deduced. This kind of activity might increase long-term retention (Hyde & Jenkins, 1969). These considerations lead to the following predictions. If subjects in the FACT condition deduce most of the inferences correctly, they should perform at least as well on the inference test as subjects in the FACT-AND-INFERENCE condition. If subjects in the FACT condition do not deduce most of the inferences correctly, subjects in the FACT-AND-INFERENCE condition should perform better.

EXPERIMENT 9

Method

Subjects. Thirty-six UCLA undergraduates participated in the two-hour experiment. Subjects were either paid $6.00 or given course credit for their participation.

Materials. The four pairs of meaningful stories about the mythical country of Morinthia used in Experiment 7 were used again in this experiment.

Each pair of stories included four pairs of related facts, such as (5) and (6) on p.133. Each pair of related facts contained sufficient information to deduce a particular inference not explicitly stated in either story (such as fact (4) above).

Design. All subjects received all four pairs of stories in an order counterbalanced across subjects. To insure high retention, subjects received three copies of each story, alternating between Story 1 and Story 2 of that pair. Subjects were randomly assigned to one of the following experimental conditions.

CONTROL Condition. CONTROL subjects were told (via printed instructions in their test booklets) that the experiment concerned their ability to read, understand, and remember prose; that they would read a number of stories about an imaginary country; and that they would later be asked questions about the stories. The subjects were further instructed to try to combine facts from the stories and to deduce inferences from combinations of facts. They were told to write down any inferences they deduced in a space provided below the stories. An example (unrelated to any of the experimental stories) was presented, showing two stories with information suitable for deducing inferences. Subjects were told that deducing inferences would maximize their performance on the tests to come.

FACT Condition. FACT subjects received the same instructions and test materials as the control subjects, with the following additions. The second story of each pair was modified by placing a numerical footnote after each fact in the story that was involved in one of the experimental inferences. This footnote referred the subject to a "hint" at the bottom of the page, which contained the related fact from the first story in the pair. Each of these pairs of facts provided the basis for deducing an inference. For example, the fact "King Egbert was a dictator," which occurred in the first story, was repeated as a footnote to the fact "Albert hated all dictators" in the second story. Subjects were told to use these hints to deduce inferences. They were also told that other inferences might be possible, so they should not restrict their efforts to facts associated with hints. Again, subjects were told to write down all inferences they deduced in a space provided on the page.

FACT-AND-INFERENCE Condition. FACT-AND-INFERENCE subjects received the same materials as FACT subjects, with the following modifications. The second sentence of each pair was again marked with a footnote, but the hint indicated by the footnote was the related fact from the first story in the pair and the inference that could be generated from the pair of related facts. For example, the footnote to "Albert hated all dictators" was "King Egbert was a dictator" and "Therefore, Albert hated King Egbert." Subjects were told that these hints would show them examples of inferences deducible from the two stories and that they should learn these. They were also told to deduce as many other inferences as possible and write them down in the space provided.

Procedure. Subjects were tested in groups. Each subject was given a booklet containing the experimental stories and tests. Subjects' progress through the booklets was self-paced. Intentional learning instructions were given, including the warning that an inference test would be given. Subjects recorded the times at which they began and finished reading each pair of stories.

After subjects had read all four pairs of stories, they were given combined recognition-verification tests. Three kinds of items were tested: OLD sentences, which had previously appeared in the stories; NEW sentences, which were made up of invalid combinations of facts from the stories; and the INFERENCES described above. Subjects were instructed to use a response line below each sentence to answer two questions: (1) Did this exact sentence appear in a story you have read (OLD or NEW)? and (2) If not, is the sentence logically true with respect to the facts presented in the stories (true or false)? Subjects were given both OLD and NEW-TRUE response options to avoid confusion over whether they were being tested for recognition or verification. The two kinds of responses were treated identically in the analysis.

There were twelve test items (four of each of the three types of items) for each set of stories. The test items were blocked by story pair, and these blocks were presented in the same order as the story pairs to control for any short-term retention of the stories. Subjects were not allowed to look back at the stories at any time during the testing.

Following the recognition-verification test, subjects verified the sixteen true and false inferences in syllogism form. That is, each inference was preceded by the two facts from the stories that presumably supported it. Subjects simply indicated whether each inference followed logically from the associated premise sentences.

Results and Discussion

Inference Deduction. The inferences deduced during learning in the CONTROL and FACT conditions were examined for the sixteen inferences around which the stories were designed. (These will be referred to as the *critical* inferences of the experiment.) FACT subjects deduced more of these critical inferences than did CONTROL subjects (probability of generation: CONTROL = .40, FACT = .84; $t(22) = 5.14$, p .01), although both groups missed a significant number of critical inferences (.84 1.0, $t(11) = 3.73$, $p < .001$). Thus, while CONTROL subjects deduced some of the inferences on their own, FACT subjects, aided by footnotes referring to previously read related facts, deduced many more of the inferences.

Recognition-Verification Test. The variable of interest is the probability of a correct response to each of the three kinds of test items. For NEW items, the correct response was NEW-FALSE. For OLD and INFERENCE items, the correct response was "true" (either OLD or NEW-TRUE). While there are many interesting questions concerning subjects' abilities to discriminate sentences they have actually seen before from valid new sentences (Bransford & Franks, 1971), our concern here is with whether or not subjects think a sentence is true, regardless of their reasons.

Failure to judge an INFERENCE to be "true" could be due to either of two causes. First, subjects may have failed to integrate the related facts underlying the inference. Alternatively, they may have integrated the facts but decided that the inference did not follow from them. Since we are primarily concerned with subjects' ability to integrate facts and not their ability to reason, performance on the syllogism task was used to control performance on the inference task. Any inference whose corresponding syllogism was incorrect was excluded from the analysis. Thus, the results below reflect performance on only those inferences that the subject believed to be logically valid when all the necessary information was readily available. Subjects in the three groups performed comparably well on the syllogisms (p(correct true syllogism): CONTROL = .92, FACT = .92, FACT-AND-INFERENCE = .96; F(2,33) = 1.27, MSe = .006; p(correct false syllogism): CONTROL = .80, FACT = .77, FACT-AND-INFERENCE = .68; F(2,33) = .82, MSe = .060, so this conditionalizing did not lead to any item selection effects.

The results are shown in Table 6.1. The probability of a correct response was tested in a group (CONTROL/FACT/FACT-AND-INFERENCE) by item (OLD/NEW/INFERENCE) analysis of variance. The main effects of group ($F(2,33) = 8.87$, $p < .001$) and item ($F(2,66) = 240.69$, $p < .001$) were significant. In addition, the interaction between group and item was significant, $F(4,66) = 5.05$, $p < .001$. All three experimental groups called OLD items "true" equally often ($F(2,33) < 1.0$). However, Newman-Keuls tests (Winer, 1962) found that subjects in the FACT and FACT-AND-INFERENCE conditions verified inferences correctly more often than did CONTROL subjects ($p < .01$). In addition, FACT

Table 6.1

PROBABILITY CORRECT FOR OLD, NEW, AND INFERENCE ITEMS

	CONTROL	FACT	FACT and INFERENCE
p (OLD or TRUE¦OLD)	.96	.96	.96
p(old or TRUE¦INFERENCE)	.80	.96	.97
p(NEW and FALSE¦NEW)	.49	.63	.54

subjects correctly classified NEW sentences more often than did CONTROL or FACT-AND-INFERENCE subjects ($p < .05$).

These results indicate that the FACT condition produced better overall performance than either the CONTROL or the FACT-AND-INFERENCE condition. Subjects in the FACT condition verified OLD sentences as accurately as subjects in either of the other two conditions. They verified true inferences more accurately than subjects in the CONTROL condition, and they verified false NEW sentences more accurately than subjects in either of the other two conditions.

The finding that subjects in the FACT condition performed as well as subjects in the FACT-AND-INFERENCE condition on the true inferences is a little surprising in light of the predictions above. We predicted that performance would be better in the FACT-AND-INFERENCE condition unless subjects in the FACT condition correctly deduced most of the test inferences during study. As noted above, subjects in the FACT condition correctly deduced 84 percent of the test inferences during study. Apparently, the active processing required to deduce inferences in the FACT condition provided a relatively large advantage. This advantage was sufficient to offset subjects' failure to deduce all of the inferences, permitting them to verify inferences as accurately as subjects in the FACT-AND-INFERENCE condition.

The importance of deducing inferences during study can be seen by conditionalizing the probability of judging an inference to be true on whether or not that inference was deduced during study (see Table 6.2). For both the CONTROL and FACT conditions, the probability of judging a previously deduced inference true was close to 1.0, while the probability of judging other inferences true was around .70 ($F(1,17) = 19.99$, $p < .001$). Neither the group effect nor the group-by-deduction interaction was significant (both F's < 1.0). Thus, the difference between the overall performance by CONTROL and FACT subjects is due to the failure of CONTROL subjects to deduce as many inferences during study. CONTROL and FACT subjects' performance on previously deduced inferences did not differ from FACT-AND-INFERENCE subjects' overall performance ($F(2,28) < 1.0$), indicating that subjects' memory for deduced inferences was at least as good as that for explicitly read inferences.

Table 6.2

PROBABILITY CORRECT FOR INFERENCES CONDITIONALIZED ON PRIOR DEDUCTION

	CONTROL	FACT	FACT and INFERENCE
Produced	.96	.98	.97
Not produced	.66	.75	...

We must consider an alternative hypothesis for CONTROL subjects' relatively poor performance on true inferences. Subjects in the CONTROL condition were given no information at all about which inferences would be tested. Subjects in the FACT-AND-INFERENCE condition read the test inferences, and subjects in the FACT condition had their attention focused on facts underlying the test inferences. Perhaps subjects in the CONTROL condition actually deduced as many correct inferences as subjects in the other conditions but did not happen to deduce those subsequently tested. Thus, subjects in the CONTROL condition may not have been given an opportunity to display their knowledge.

In order to test this alternative hypothesis, we scored subjects' performance for the number of non-critical inferences deduced during study. The number of non-critical inferences deduced decreased as the amount of information given in footnotes increased (number of non-critical inferences deduced: CONTROL = 15.75, FACT = 13.00, FACT-AND-INFERENCE = 10.42), but this effect was not significant (F(2,33) = 1.09, MSe = 78.10). Thus, we cannot explain subjects' poor performance in the CONTROL condition as a consequence of their having generated the "wrong" inferences.

Reading Times. The time to read and process pairs of stories averaged 1.9 minutes in the CONTROL condition, 2.3 minutes in the FACT condition, and 2.4 minutes in the FACT-AND-INFERENCE condition. The effect of condition was significant ($F(2,33) = 3.21$, $p < .05$), indicating that subjects in the two annotation conditions spent more time processing the stories than did those in the CONTROL condition. This is not surprising, given the different activities performed in the three conditions: Subjects in the CONTROL condition read only the two stories and deduced an average of 22.2 inferences (critical and non-critical). Subjects in the FACT condition read the two stories plus four facts repeated in footnotes and deduced an average of 25.8 inferences. Subjects in the FACT-AND-INFERENCE condition read the two stories plus four repeated facts and associated inferences in footnotes and deduced an average of 10.4 additional inferences.

GENERAL DISCUSSION

These results suggest that active efforts to integrate studied material and deduce valid inferences are valuable study behaviors. Our subjects' knowledge of previously deduced inferences was nearly perfect, while their knowledge of other inferences was only slightly above chance.

Simply instructing subjects to attempt to integrate related information from studied texts and illustrating how to deduce inferences apparently can improve performance. Our control subjects successfully deduced 40 percent of the critical inferences without any hints at all and, as a consequence, verified 80 percent of the test inferences correctly. This represented a substantial improvement over the performance observed in previous studies in which subjects were given more general learning instructions.

The simple annotation techniques illustrated in the present experiment apparently can further improve subjects' integration of related information and deduction of valid inferences. While subjects take more time to study annotated stories, we cannot expect to improve performance at no cost. The relatively small investment of another 20 or 30 seconds subjects made in the annotation conditions paid off in substantially improved performance.

Apparently, the major impediment to successful performance is failure to detect the relationships among separately encountered facts, rather than the inability to reason correctly. Thus, it appears that simply pointing out the relationships among separately encountered facts (the FACT condition) is as effective as deducing the inferences for the subjects (the FACT-AND-INFERENCE condition). The finding that subjects correctly rejected NEW sentences more often in the former condition than in the latter suggests that the simpler technique may actually produce the best overall performance.

VII. CONCLUSIONS

The studies described in this report explored a broad range of issues in human processing of knowledge from text. These issues include the following:

1. Which information people acquire during incidental and intentional learning and how they acquire it.
2. Transfer effects in learning from related texts.
3. Conditions under which acquired information can be recognized or retrieved.
4. Integration of related facts encountered in the same or different texts.
5. Inferencing based on related facts encountered in different texts.
6. Searching an external text or memory of a text for information relevant to a hypothetical inference.
7. Annotation techniques for facilitating inferential reasoning and retention of inferences from information in studied texts.

Throughout this report, we have assumed a single, unifying theoretical framework. We used this theory to motivate the particular experiments we conducted and to account for the results we obtained. This theory embodied principles governing acquisition, transfer, recognition, and retrieval of information from text, as well as specifying the details of the memory representations on which these processes operate. Based on our experimental results, we may now summarize the theoretical assumptions that have gained empirical support.

A knowledge structure that represents information from one context can be used to represent the same information occurring in different contexts. The information shared across contexts is represented as a configuration of general, or variable, concepts and their relations (a schema). The details that specify or "instantiate" the variable concepts in each individual context are associated with their respective concepts by context-preserving relations. When a detail is associated with a well-learned schema, the schema must be retrieved in order to retrieve the detail that instantiates it.

The use of a schema for encoding or retrieving information depends on its accessibility in memory; that is, the probability that it can be activated, either for use in storage of incoming information or for retrieval of previously stored information. Accessibility depends upon the strength of the stored information, the extent of the overlap or match between input and schema, and the recency of previous activations. Each time a schema is activated for use, it becomes more accessible for successive activations.

When multiple details instantiate a variable concept in a schema, they compete with one another for associations with the variable concept. As the number of competing details increases, a person's ability to discriminate (and thus recall correctly) the context in which each detail occurred decreases. Thus, when a schema is used to encode multiple complementary facts (as in Chap. IV), memory is enhanced. However, when a schema is used to encode multiple competing facts (as in Chap. II), memory can be interfered with.

Schemata may also be viewed as more global knowledge structures that encode entire texts. At this level of analysis, a schema provides a framework for sequencing and organizing the events in a narrative discourse.

The main purpose of the studies described in this report was to discover techniques for improving the amount of information people can assimilate from texts and their ability to use that information when necessary. Based on these studies, a set of principles for improving human learning and performance with texts has emerged. These principles are listed below, accompanied by brief descriptions of supporting data from the particular experiments from which the principles were derived. We believe that these principles could serve as useful guidelines in designing information systems of the future.

1. Presentation of new information in well-learned structural organizations can facilitate learning of that information.

It is frequently necessary to present texts containing related facts--that is, facts having the same general form, but different details. For example, one might be required to learn a series of profiles of individuals in which the fact "He was born in the year ..." was common to each profile, but the actual year of birth was different for each person. In Experiment 1 (Chap. II), subjects' memory for both the general form of a fact (e.g., someone was born in some year) and the details (e.g., the actual year of birth) improved by up to 100 percent when the text in which the fact appeared was preceded by one, two, or three different texts containing related facts. Thus, presenting a series of related facts in well-learned structures is desirable because it facilitates memory for the related facts.

2. Blocking presentation of large numbers (on the order of five or more) of texts containing related facts interferes with learning.

Experiment 1 also showed that immediately preceding a text by five or more texts containing related facts further improved memory for the general forms of related facts. However, memory for the details of related facts deteriorated by up to 50 percent. This occurred because subjects had difficulty remembering which detail (e.g., year of birth) went with which text. Thus, blocking presentation of large numbers of texts containing related facts should be avoided because it interferes with memory for the details of related facts.

3. Temporal separation in presentation of related texts can preserve facilitating effects and eliminate interfering effects.

Whenever related texts must be read, there is a good chance that memory for the information in the texts will suffer interference. For example, Experiment 1 showed that interference would occur if presentations of large numbers of texts containing related facts were blocked. However, in Experiment 2, it was also found that interference effects could be eliminated and memory for details specific to individual texts could be improved by up to 150 percent by inserting a reasonable temporal interval, such as 24 hours, between presentations of related texts. Thus, presentation of potentially interfering related texts should be temporally separated to preclude interference effects and to facilitate memory for the details of studied texts.

4. Elimination of redundancy and irrelevant commentary from newspaper stories facilitates assimilation and retention of important facts.

Newspaper stories are written in accordance with stylistic conventions of journalism. The most timely, important, or eye-catching information is placed in the opening sentences of a story, and the details and background information are elaborated in subsequent paragraphs. This organization results in distortions of the true narrative sequence of events, redundant repetition of facts, and commentary that is tangential to and less important than the newsworthy events. In Experiments 3 and 4 (Chap. III), newspaper stories were restructured by grouping related information together either in a sequential narrative or by primary topic, and by eliminating redundant or unimportant information. Retention of important information was facilitated by all reorganizations of news story facts that deleted the irrelevant information. In addition, the redundant and unimportant information was rarely recalled from the news stories. This indicates that when redundant and irrelevant information is present in stories, it is rarely learned by readers and it inhibits acquisition of more salient facts.

5. Text organizations that place complementary facts in close proximity improve integration of those facts.

Frequently, texts contain complementary information. The complementarity between separately occurring facts permits them to be integrated into a single, composite fact. For example, one fact might specify that George Washington was the first President of the United States and another might specify that he lived at Mount Vernon. It is important for the reader to integrate these facts in order to realize that the first President of the United States lived at Mount Vernon. In Experiments 5 and 6 (Chap. IV), presenting complementary facts in succession within a text, rather than separating them with other related facts, improved performance by up to 100 percent. Thus, text organizations that maximize the proximity of complementary facts are desirable because they facilitate integration of those facts.

6. Wording complementary texts as similarly as possible improves integration of complementary facts that occur in separate texts.

It is frequently impossible to organize texts so that complementary facts occur together or even in the same text. A text may contain information that complements information from a prior text. For example, the fact that George Washington was the first President of the United States might occur in an American History text, while the fact that he lived at Mount Vernon might occur in a biographical text. It is important for the reader to be able to integrate complementary facts even though they occur in separate texts. Experiments 5 and 6 showed that wording complementary facts that occur in separate texts as similarly as possible improved integration of the facts by up to 100 percent. Thus, it is desirable to word complementary texts as similarly as possible.

7. Wording related texts as similarly as possible improves inferential reasoning based on facts within the texts.

Sometimes, the information in a text is tangentially related to information that occurred in a previously learned text. While the relationship may not lead to simple integration of related facts into a composite fact, it may provide a basis for inferential reasoning. For example, a text covering Early American History might specify that George Washington was the first President of the United States. Another text covering Modern American History might specify that Franklin Delano Roosevelt was the first U.S. President to serve more than two terms of office. Using these two facts, the reader could infer that George Washington served no more than two terms of office. Experiment 7 (Chap. IV) showed that similar wording of tangentially related texts occurring in separate texts improved inferential reasoning based on those facts by up to 50 percent. Thus, related texts should be worded as similarly as possible to facilitate inferential reasoning based on the information in the texts.

8. Reasoning from memory of carefully studied texts is more accurate than reasoning based on inspection of less familiar texts.

People frequently have at their disposal a number of documents that provide the information on which decisions are to be based. A standard procedure is to use such documents as reference sources, searching them for particular facts or categories of information as they are needed. Experiment 8 (Chap. V) indicates that this is an effective strategy only if the information needed is present in literal form in the available documents. It is an ineffective strategy if the information needed is available only as an inference based on a configuration of facts that occur separately in the documents. In the latter case, people perform poorly at obtaining necessary information. However, their performance can be improved by at least 50 percent if they study to learn the reference documents before attempting to use them. Thus, having the reader study reference texts is desirable because it facilitates the ability to detect important relationships among separately occurring facts.

9. Studying to learn texts improves knowledge of the information the texts contain over using the texts to perform inferencing.

As discussed above, people frequently fail to detect important relationships among facts if they have not first studied the documents in which the facts occur. In addition, Experiment 8 showed that people learn very little about information contained in a text unless they consciously try to learn. For example, although people spend a considerable amount of time inspecting available texts in their (usually unsuccessful) attempts to retrieve related facts, they learn almost nothing. In contrast, people learn a great deal if they spend a comparable amount of time studying the available texts with the intention of learning the information they contain. Thus, the reader ought to study reference texts not only to facilitate detection of configurations of related facts, but also to facilitate future retrieval of individual facts.

10. Annotating texts with references to related facts that have occurred in previous texts facilitates general inferential reasoning from the texts.

Experiments 5 to 7, discussed above, indicated that people have difficulty integrating complementary facts and drawing inferences from related facts in separate texts. However, it is not always possible to organize texts so that complementary and related facts occur together. In Experiment 9, a relatively simple annotation method was found to improve inferential reasoning based on facts occurring in separate texts by up to 100 percent. This annotation method involved repeating previously read, related facts as footnotes to the appropriate facts in a text. It was neither necessary nor desirable to generate the appropriate inference for the reader and include it in the footnote. This annotation method did not improve performance on inferences drawn from the annotated facts. However, it appeared to inhibit generation of other inferences, in addition to those based on the annotated facts. Thus, texts should be annotated with references to prior, related facts in order to facilitate reasoning from the texts.

REFERENCES

Anderson, J. R. Memory for information about individuals. *Memory and Cognition,* 1977, *5*, 430-442.

Anderson, J. R. *Language, memory and thought.* Hillsdale, N.J.: Erlbaum Assoc., 1976.

Anderson, J. R. and Bower, G. H. *Human associative memory.* Washington, D.C.: V.H. Winston, 1973.

Anderson, J. R., and Hastie, R. Individuation and reference in memory: Proper names and definite descriptions. *Cognitive Psychology,* 1974, *6*, 495-514.

Anderson, R. C. The notion of schemata and the educational enterprise. In R. C. Anderson, R. J. Spiro and W. E. Montague (eds.), *Schooling and the acquisition of knowledge.* Hillsdale, N.J.: Erlbaum, 1977.

Anderson, R. C., and Biddle, W. B. On asking people questions about what they are reading. In G. H. Bower (ed.), *Psychology of learning and motivation* (Vol. 9). New York: Academic Press, 1975.

Anderson, R. C., and Myrow, D. L. Retroactive inhibition of meaningful discourse. *Journal of Educational Psychology,* 1971, *62*, 81-94.

Anderson, R. C., et al. Frameworks for comprehending discourse. Technical Report 12, Laboratory for Cognitive Studies in Education, University of Illinois, 1976.

Anderson, R. C., Spiro, R. J., and Anderson, M. C. Schemata as scaffolding for the representation of information in connected discourse. Technical Report 24, Center for the Study of Reading, University of Illinois, 1977.

Barclay, J. R. The role of comprehension in remembering sentences. *Cognitive Psychology,* 1973, *4*, 229-254.

Bartlett, F. *Remembering.* Cambridge: Cambridge Press, 1932.

Bobrow, D. G., and Norman, D. A. Some principles of memory schemata. In D. G. Bobrow and A. Collins (eds.), *Representation and understanding.* New York: Academic Press, 1975.

Bower, G. H. Selective facilitation and interference in retention of prose. *Journal of Educational Psychology,* 1974, *66*, 1-8.

Bransford, J., Barclay, J., and Franks, J. Sentence memory: A constructive versus interpretive approach. *Cognitive Psychology,* 1972, *3*, 192-209.

Bransford, J. and Franks, J. The abstraction of linguistic ideas. Cognitive Psychology, 1971, 2, 331-350.

Clark, H. and Haviland, S. Comprehension and the given-new contract. In R. Freedle (ed.), Discourse production and comprehension. Norwood, N.J.: Ablex, 1977.

Craik, F.I.M., and Lockhart, R. S. Levels of processing: A framework for memory research. Journal of Verbal Learning and Verbal Behavior, 1972, 11, 671-684.

Crouse, J. H. Retroactive interference in reading prose materials. Journal of Educational Psychology, 1971, 62, 39-44.

van Dijk, T., and Kintsch, W. Cognitive psychology and discourse: Recalling and summarizing stories. In W. Pressler (ed.), Trends in text-linguistics. New York: deGruyter, 1977.

Dooling, D. J., Christiaansen, R. E., and Keenan, T. F. Levels of encoding and the forgetting of prose. Paper presented at the meeting of the Psychonomic Society, Denver, November, 1975.

Fillenbaum, S. Memory for gist: Some relevant variables. Language and Speech, 1966, 9, 217-227.

Foos, P. W., Smith, K. H., Sabol, M. A., and Mynatt, B. T. Constructive processes in simple linear-order problems. Journal of Experimental Psychology: Learning and Memory, 1976, 2, 759-76.

Frase, L. T. Prose processing. In G. H. Bower (ed.), Psychology of learning and motivation (Vol. 9). New York: Academic Press, 1975.

Frase, L. Integration of written text. Journal of Educational Psychology, 1973, 65, 252-261.

Frase, L. T. Structural analysis of the knowledge that results from thinking about text. Journal of Educational Psychology Monograph, 1969, 60. (a)

Frase, L. Paragraph organization of written materials: The influence of conceptual clustering upon level of organization. Journal of Educational Psychology, 1969, 60, 394-401. (b)

Frase, L. T., and Schwartz, B. J. Effect of question production and answering on prose recall. Journal of Educational Psychology, 1975, 67, 628-635.

Frederiksen, C. Effects of context-induced processing operations on semantic information acquired from discourse. Cognitive Psychology, 1975, 7, 139-166. (a)

Frederiksen, C. Representing logical and semantic structure of knowledge acquired from discourse. Cognitive Psychology, 1975, 7, 371-458. (b)

Friedman, M. and Greitzer, F. Organization and study time in learning from reading. Journal of Educational Psychology, 1972, 63, 609-616.

Gagne, E. and Rothkopf, E. Text organization and learning goals. Journal of Educational Psychology, 1975, 67, 445-450.

Garrod, S. and Sanford, A. Interpreting anaphoric relations: The integration of semantic information while reading. Journal of Verbal Learning and Verbal Behavior, 1977, 16, 77-90.

Garrod, S., and Sanford, A. Anaphora: A problem in text comprehension. Proceedings of the Psychology of Language Conference, Stirling, Scotland, 1976.

Garrod, S. and Sanford, A. Anaphora: A problem in text comprehension. To appear in Proceedings of the Psychology of Language Conference, Stirling, Scotland, 1976.

Green, D. M, and Swets, J. A. Signal detection theory and psychophysics. Huntington, New York: Krieger Publshing Co., 1966.

Haviland, S. and Clark, H. What's new? Acquiring new information as a process in comprehension. Journal of Verbal Learning and Verbal Behavior, 1974, 13, 512-521.

Hayes-Roth, B. Evolution of cognitive structures and processes. Psychological Review, 1977, 84, 260-278.

Hayes-Roth, B. and Hayes-Roth, F. The prominence of lexical information in memory representations of meaning. Journal of Verbal Learning and Verbal Behavior, 1977, 16, 119-136.

Hayes-Roth, B. and Hayes-Roth, F. Plasticity in memorial networks. Journal of Verbal Learning and Verbal Behavior, 1975, 14, 506-522.

Head, H. Studies in neurology. New York: Oxford University Press, 1920.

Hebb, D. O. Organization of behavior. New York: John Wiley & Sons Inc., 1949.

Hupet, M. and LeBouedec, B. The given-new contract and the constructive aspect of memory for ideas. Journal of Verbal Learning and Verbal Behavior, 1977, 16, 69-75.

Hyde, T. S., and Jenkins, J. J. Differential effects of incidental tasks on the organization of recall of a list of highly associated words. Journal of Experimental Psychology, 1969, 82, 472-481.

James, C. T., Hillinger, M. L., and Murphy, B. J. The effects of complexity on confidence ratings in linguistic integration. Memory & Cognition, 1977, 5, 355-361.

Katz, S., Ateson, B., and Lee, J. The Bransford-Franks linear effect: Integration or artifact? Memory & Cognition, 1974, 2, 709-713.

Katz, S. and Gruenewald, P. The abstraction of linguistic ideas in "meaningless" sentences. Memory & Cognition, 1974, 2, 737-741.

Kintsch, W. Memory for prose. In C. N. Cofer (ed.), The structure of human memory. San Francisco: Freeman, 1975.

Kintsch, W. The representation of meaning in memory. Hillsdale, N.J.: Lawrence Erlbaum Assoc., 1974.

Kintsch, W., Mandel, T. and Kozminsky, E. Summarizing scrambled stories. Memory and Cognition, 1977, 5, 547-552.

Kozminsky, E. Altering comprehension: The effect of biasing titles on text comprehension. Memory and Cognition, 1977, 5, 482-490.

Kuhara, K. Relatedness and retention of prose. Japanese Psychological Research, 1976, 18, 191-195.

Kuipers, B. J. A frame for frames: Representing knowledge for recognition. In D. G. Bobrow & A. Collins (eds.), Representation and understanding. New York: Academic Press, 1975.

Kulhavy, R. W., Schmid, R. F., and Walker, C. H. Temporal organization in prose. American Educational Research Journal, 1977, 14, 115-123.

Mandler, J. and Johnson, N. Remembrance of things parsed: story structure and recall. Cognitive Psychology, 1977, 9, 111-151.

Meyer, B. The organization of prose and its effect on memory. Amsterdam: North-Holland Publishing Co., 1975.

Miller, G. A. The magical number seven, plus or minus two: Some limits on our capacity for processing. Psychological Review, 1956, 63, 81-96.

Minsky, M. A framework for the representation of knowledge. In P. H. Winston (ed.), The psychology of computer vision. New York: McGraw-Hill, 1975.

Moore, J., and Newell, A. How can MERLIN understand? In L. W. Gregg (ed.), Knowledge and cognition. Potomac, Maryland: Lawrence Erlbaum Assoc., 1974.

Myers, J., Pezdek, K., and Coulson, D. Effects of prose organization upon recall. Journal of Educational Psychology, 1973, 65, 313-320.

Norman, D., and Bobrow, D. On the role of active memory processes in perception and cognition. In C. Cofer (ed.), The structure of human memory. San Francisco: Freeman, 1975.

Palmer, S. E. Visual perception and world knowledge: Notes on a model of sensory-cognitive interaction. In D. A. Norman, D. E. Rumelhart, and the LNR Research Group, Explorations in cognition. San Francisco: Freeman, 1975.

Park, D.C. and Whitten, W.B. Abstraction of linguistic, imaginal, and pictorial ideas. Journal of Experimental Psychology, 1977, 3, 525-538.

Perlmutter, J., and Royer, J. M. Organization of prose materials: Stimulus, storage, and retrieval. Canadian Journal of Psychology, 1973, 27, 200-209.

Peterson, R. G. and McIntyre, C. W. The influence of semantic "relatedness" on linguistic integration and retention. American Journal of Psychology, 1973, 86, 697-706.

Pichert, J. W. and Anderson, R. C. Taking different perspectives on a story. Journal of Educational Psychology, 1977, 69, 309-315.

Potts, G. R. Integrating new and old information. Journal of Verbal Learning and Verbal Behavior, 1977, 16, 305-320.

Potts, G. R. Information processing strategies used in the encoding of linear orderings. Journal of Verbal Learning and Verbal Behavior, 1972, 11, 727-740.

Royer, J. M., and Cable, G. W. Facilitated learning in connected discourse. Journal of Educational Psychology, 1975, 67, 116-123.

Royer, J. M., and Cable, G. W. Illustrations, analogies, and facilitative transfer in prose learning. Journal of Educational Psychology, 1976, 68, 205-209.

Rumelhart, D. Notes on a schema for stories. In D. Bobrow and A. Collins (eds.), Representation and understanding: studies in cognitive science. New York: Academic Press, 1975.

Rumelhart, D. E., Lindsay, P. H., Norman, D. A. A process model for long-term memory. Organization of memory. New York: Academic Press, 1972.

Rumelhart, D. E. and Ortony, A. The representation of knowledge in memory. In R. C. Anderson, R. J. Spiro, and W. E. Montague (eds.), Schooling and the acquisition of knowledge. Hillsdale, N. J.: Erlbaum, 1977.

Sachs, J. S. Memory in reading and listening to discourse. Memory and Cognition, 1974, 2, 95-100.

Sasson, R. Semantic organizations and memory for related sentences. The American Journal of Psychology, 1971, 84, 253-267.

Schank, R. The role of memory in language processing. In C. Cofer (ed.), The structure of human memory. San Francisco: Freeman, 1976.

Schank, R. C. The structure of episodes in memory. In D. G. Bobrow A. Collins (eds.), Representation and understanding. New York: Academic Press, 1975.

Schank, R., and Abelson, R. Scripts, plans, goals, and understanding. Hillsdale, N.J.: Lawrence Erlbaum Assoc., 1977.

Schank, R., and Abelson, R. Scripts, plans and knowledge. Proceedings of the Fourth International Joint Conference on Artificial Intelligence, Tbilisi, USSR, 1975, 151-157.

Schultz, C. and DiVesta, F. Effects of passage organization and note taking on the selection of clustering strategies and on recall of textual materials. Journal of Educational Psychology, 1972, 63, 244-252.

Thorndyke, P. W. Pattern-directed processing of knowledge from texts. In D. A. Waterman and F. Hayes-Roth (eds.), Pattern-Directed Inference Systems. New York: Academic Press, 1978.

Thorndyke, P. W. Cognitive structures in comprehension and memory of narrative discourse. Cognitive Psychology, 1977, 9, 77-110.

Thorndyke, P. W. The role of inferences in discourse comprehension. Journal of Verbal Learning and Verbal Behavior, 1976, 15, 437-446.

Underwood, B. J. The effect of successive interpolations on retroactive and proactive inhibition. Psychological Monographs, 1945, 59.

Winer, B. J. Statistical principles in experimental design. New York: McGraw-Hill, 1962.

Winograd, T. Frame representations and the declarative-procedural controversy. In D. G. Bobrow and A. Collins (eds.), Representation and understanding. New York: Academic Press, 1975.

Winograd, T. A framework for understanding discourse. In P. Carpenter and M. Just (eds.), Cognitive processes in comprehension. Hillsdale, N.J.: Lawrence Erlbaum Assoc., 1977.

Woodworth, R. S. Experimental psychology. New York: Holt and Company, 1938.

Yekovich, F.R. and Walker, C.H. Identifying and using referents in sentence comprehension. Journal of Verbal Learning and Verbal Behavior, (in press).